EXPLORING PROTEIN-RNA INTERACTIONS WITH SITE-DIRECTED MUTAGENESIS AND PHAGE DISPLAY

BY

YAN FAN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2009

Urbana, Illinois

Doctoral Committee:

Associate Professor Anne M. Baranger, Chair
Professor John A. Katzenellenbogen
Professor Huimin Zhao
Associate Professor Scott K. Silverman

ABSTRACT

The importance of RNA as a mediator of genetic information is widely appreciated. RNA molecules also participate in the regulation of various post-transcriptional activities, such as mRNA splicing, editing, RNA stability and transport. Their regulatory roles for these activities are highly dependent on finely tuned associations with cognate proteins. The RNA recognition motif (RRM) is an ancient RNA binding module that participates in hundreds of essential activities where specific RNA recognition is required. We have applied phage display and site-directed mutagenesis to dissect principles of RRM-controlled RNA recognition. The model systems we are investigating are U1A and CUG-BP1. In this dissertation, the molecular basis of the binding affinity of U1A-RNA beyond individual contacts was investigated. We have identified and evaluated the contributions of the local cooperativity formed by three neighboring residues (Asn15, Asn16 and Glu19) to the stability of the U1A-RNA complex. The localized cooperative network was mapped by double-mutant cycles and explored using phage display. We also showed that a cluster of these residues forms a "hot spot" on the surface of U1A; a single substitution at position 19 with Gln or His can alter the binding properties of U1A to recognize a non-cognate G4U RNA. Finally, we applied a deletion analysis of CUG-BP1 to define the contributions of individual RRMs and RRM combinations to the stability of the complex formed between CUG-BP1 and the GRE sequence. The preliminary results showed RRM3 of CUG-BP1 is a key domain for RNA binding. It possibly binds to the GRE sequence cooperatively with RRM2 of CUG-BP1. RRM1 of CUG-BP1 is not required

for GRE recognition, but may be important for maintaining the stability of the full-length CUG-BP1.

ACKNOWLEDGMENTS

I would like to express the deepest appreciation to my advisor, Professor Anne M. Baranger, for her guidance and support with this project. I also wish to thank my committee members, Professor John A. Katzenellenbogen, Professor Huimin Zhao, and Professor Scott K. Silverman, who offered guidance and assistance. And Finally, I would like to thank my wife, parents, sisters and brothers in China, for their endless love, support and patience.

TABLE OF CONTENTS

CHAPTER 1

RNA RECOGNITION BY RNA-BINDING MOTIF

1.1 Introduction

RNA plays numerous functions in different cellular processes (*1-5*). Besides its role as a carrier for genetic information, it also functions as a regulator of many essential biological functions, including pre-mRNA splicing, editing, RNA transport and stability (*3-4, 6*). In contrast to DNA, which exists as a double-stranded helical structure, RNA is a single-stranded molecule in most of its biological roles. It folds to form complex tertiary structures with different secondary structures embedded, such as double-stranded stem regions (A-form), internal loops, bulges and hairpins (*7*).

In eukaryotic cells, RNA never exists on its own, but always associates with different RNA-binding proteins, often forming stable ribonucleoprotein complexes (RNP) (*8-9*). The great structural diversity of RNA and the large variety of molecular functions regulated by RNA require a wide range of RNA-binding proteins. Abnormalities and disruption of protein-RNA interactions often lead to diseases (*8, 10*). Therefore, a better understanding of the molecular basis underlying RNA-protein recognitions is critical to decipher the associated cellular functions and possibly design tools to regulate RNA-protein complexes in fundamental and biomedical research.

1.2 Conserved RNA-binding modules

Many different types of RNA-binding proteins have been identified in the past decade (*11-15*). Surprisingly, most of these RNA-binding proteins have modular structures (*11, 16-21*). Single or multiple copies of these modules from the same or different structural families build a variety of binding surfaces to allow specific recognition of RNA. Here,

we focus on structural and molecular analysis of how these conserved RNA-binding modules specifically recognize RNA.

1.2.1 RNA recognition motif (RRM)

The RNA recognition motif (RRM) or ribonucleoprotein domain (RNP) belongs to the α/β protein family. It is the most abundant and best-characterized RNA-binding motif, both in structure and function (*22-23*). More than 10,000 RRMs have been identified to play various roles in post-transcriptional gene-expression processes including RNA, such as capping, pre-mRNA splicing, polyadenylation and translocation (*24*).

The RRM is typically 80-100 residues long and folds into a compact globular structure, comprised of a four-stranded anti-parallel β-sheet supported by two α helices (Figure 1-1). Structural and biochemical studies have shown that RRM proteins bind primarily to single-stranded RNA sequences (*25*). The structure of the RRM is identified by two conserved short sequences, referred to as RNP1 and RNP2, in the two center strands of the β-sheet. The β-sheet of the RRM is the major binding surface for RNA recognition (*25-26*).

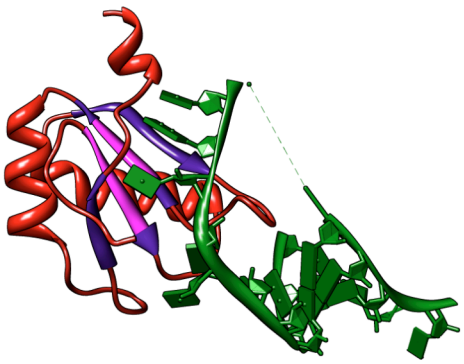


Figure 1-1. The N-terminal RRM of the U1A protein in complex with the U1 snRNA hairpin II loop. The RNP1 and RNP2 consensus regions are shown in magenta (PDB code 1URN).

Conserved charged and aromatic residues of RNP1 and RNP2 contribute to RNA recognition. Arg or Lys residues that are usually located in RNP1 can form electrostatic

interactions with the phosphodiester backbone of RNA. Phe and Tyr residues in RNP1 and RNP2 participate in stacking interactions with bases in the RNA. These conserved interactions, although crucial for RNA binding, do not contribute to the specificity of RNA recognition. A number of other well-conserved residues in this region, such as Val, Leu and Ile, are directed inside the folded domain, along with other hydrophobic residues in the two α helices, forming a hydrophobic core. These conserved residues and their orientations are crucial for constructing the major binding surface and maintaining the integrity of RRM. The RRM utilizes various loops connecting different β strands and contiguous residues in N- and C-terminal regions of RRM to make additional contacts with the RNA target. These interactions are thought to be the major determinants of RNA binding specificity (*25*).

A single RRM is sufficient to bind to specific RNA sequences in certain circumstances (*25*). Generally, a single RRM recognizes four to eight RNA nucleotides with a variety of affinities and specificities. For example, the N-terminal RRM of the U1A protein binds to seven nucleotides in the recognition element of U1 snRNA hairpin loop with an affinity of $K_D \approx 10^{-11}$ M (*27-28*). In many cases, RNA-binding proteins require multiple copies of RRMs to achieve higher affinity and specificity for RNA (*29*). These tethered RRMs create an extended versatile RNA-binding surface to accommodate a large structural diversity of RNA (*30-31*). Proteins with multiple RRM domains can recognize larger RNA sequences than can proteins with only a single RRM. Interestingly, some members of the RRM family have been found to interact with other proteins (*32-33*). These interactions could be important for regulating the binding of a given protein to RNA as well as facilitating the assembly of multiple protein complexes during RNA

regulation processes. These observations broaden the functions of this important protein family.

1.2.2 KH domain

The heterogeneous nuclear RNP K homology (KH) domain was identified in the human hnRNP K protein and is regularly found in living cells of eukaryotes, eubacteria and archaea (*34*). It binds to single-stranded RNA with different affinities and specificities (*35*). For example, the KH3 domain of hnRNP K recognizes poly-C ssRNA with a K_D of approximately 10^{-6} M (*36*). The Nova KH3 domain binds to ssRNA with much tighter affinity ($K_D \approx 10^{-9}$ M) (*37*). KH domains have been associated with several levels of the regulation of gene expression and implicated in many genetic diseases. For example, a single mutation within the highly conserved residues in FMR1 disrupts its association with RNA, causing fragile-X mental retardation syndrome (*38*). The KH domain is composed of ~70 residues and shares the same α/β fold structure as the RRM (*39-40*). It is characterized by a compact structure with three α-helices supported by a three-stranded β-sheet (Figure 1-2).

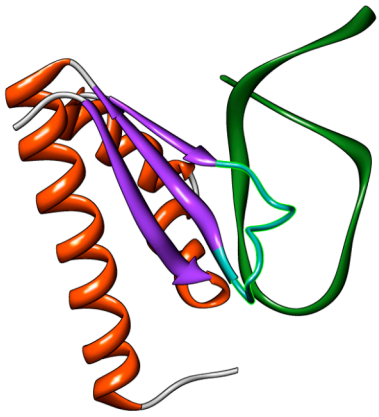


Figure 1-2. The K-homology-3 (KH3) domain of Nova-2 in complex with 5'-AUCAC-3'. A conserved GXXG sequence that is located in an exposed loop (light green) mediates binding to both single-stranded DNA and RNA (PDB code 1EC6).

The KH domain has an important signature sequence of (I/L/V)IGXXGXX(I/L/V), X represents Lys, Arg or Gly, near the center of the domain (*34, 41*). The highly conserved GXXG segment comprises the loop region that connects α helix 1 and α helix 2 in the fold. This GXXG segment always associates with the loops connecting the β strands for RNA recognition. These loops may be crucial for the high specificity since they are shown to be flexible without RNA bound in isolated domains from hnRNP K and FMR1 (*34, 38, 42*).

Like RRMs, KH domains are present as single or multiple copies in proteins to achieve independent or cooperative RNA recognition. However, the KH domain binds to RNA with a completely different pattern of interactions than the RRM. KH domains use the exposed combinations of α helices, loops and β strands to recognize RNA. Because there are no aromatic residues present in the KH domain, affinity and specificity are achieved mainly by hydrogen bonds, electrostatic interactions, van der Waals and shape complementarity (*41*).

Two different topologies, named the type I and type II folds, have been identified in the KH domain family (*39*). They differ in the arrangements of the common βααβ core. The type I fold, like KH3 of hnRNP K, has the common core at the N-terminus followed by an additional β strand and an α helix (βααββα) (*43*), while the type II fold is αββααβ, with the common core at the C-terminus. The β sheet is antiparallel in type I but mixed in type II KH folds. These specific topologies have important implications for the functions of related proteins, which can be seen from the complex of the Nova-2 KH3 domain bound to a tetranucleotide hairpin loop (*40*). In the homologous Nova proteins, Nova 1 and Nova 2 proteins contain three contiguous KH domains (KH1, KH2 and KH3) linked by various lengths of interdomain linkers, a short linker (26-50 residues) between KH1 and KH2 and a much longer linker (179-204 residues) between KH2 and KH3 (*44*). The C-terminal KH3 domains of Nova-1 and Nova-2 proteins are the key domains for recognizing tetranucleotide sequences (UCAY) in hairpin loops. In the crystal structure of Nova-2 bound to the single-stranded UCAC sequence in an RNA hairpin loop (*40*), the bases are splayed out to make contacts with the hydrophobic binding platform comprised of the two helices (α1 and α2) and the β2 strand. The ssRNA is mainly gripped by the core sequence GXXG and another variable loop. The two flanking Gly residues in the core sequence appear crucial for subtle structural adjustments that enable easy access of RNA to the backbone of the KH domain. Affinity and specificity are mediated by contacts between the two central nucleotides C and A and side chains of proximate residues in the KH3 domain, such as Glu and Arg, through specific hydrogen bonding interactions (*40*).

1.2.3 Double-stranded RNA-binding motif (dsRBM)

The dsRBM is found in proteins in viruses, bacteria and eukaryotes (*41, 45-47*). Over 100 homologous sequences have been identified by the PFAM domain alignment database (*48*). NMR structures of *Drosophila melanogaster* Staufen (*49*), *Escherichia coli* RNase III (*50*), and two dsRBMs of human PKR (*51*) revealed a common feature of this motif. It is a α/β fold (αββββα), in which the two α helices are packed against three stranded anti-parallel β-sheet. There are many conserved residues scattered throughout the 65-75 residues long protein sequences, mostly in the N- and C-terminal regions of the motif. These residues are often orientated on one side of the motif and are exposed to the solvent, suggesting they may play important roles in RNA recognition. This motif binds to A-form RNA in the minor and major grooves as shown by the structure of the *Xenopous laevis* protein Xlrbpa in complex with its cognate dsRNA (Figure 1-3) (*52*) . The RNA target is comprised of two short RNA sequences (10 base pair long for each) stacked together, which form a continuous A-form double helix. The dsRBM recognizes 16 base pairs of the RNA sequence and engages them predominantly on one face of the duplex: the N-terminal α helix (α1) and the loop between β 1 and β 2 bind to two adjacent minor grooves. Some residues in this region, such as Glu and His, are involved in direct and water-mediated hydrogen bonding interactions with the nucleotides in the minor grooves. The C-terminal α helix (α2) of the motif interacts with the major groove of RNA. The A-form duplex is partially unwound in order to widen the major groove to accommodate the C-terminal α helix. Several conserved Lys residues make direct contacts with the phosphodiester backbone and ribose 2'-OH groups of RNA to stabilize the complex (*53-56*).

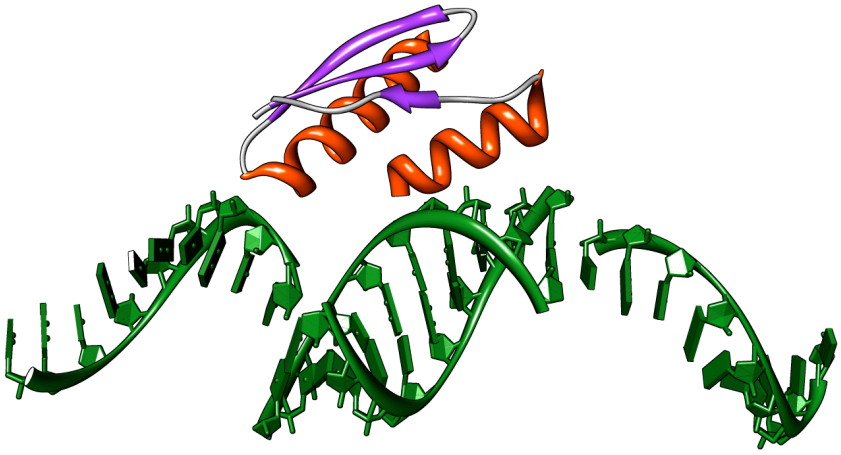


Figure 1-3. The structure of Xlrbps dsRBM2 in complex with its cognate dsRNA (PDB code 1DI2).

dsRBD binds to double-stranded RNA with low sequence specificity, and does not bind to dsDNA or DNA-RNA hybrids. In certain circumstances, specificity can partially be achieved for the recognition of irregular helical elements, such as hairpins, mismatches, and bulges (*57-60*). Many double-stranded RNA binding proteins, such as ADAR1(*54*) and Staufen (*59, 61-62*), require multiple dsRBMs for optimal RNA recognition. Cooperativity of recognition and the extended binding interfaces enable specific binding of certain RNA structures (*56, 63*). The spacing of the motifs within a protein also plays a role in the specificity of the protein for RNA. It should be noted that many proteins use different topologies other than αββββα to bind double-stranded RNA. The influenza viral protein NS1 and 2', 5'-oligoadenylate synthetase (OAS) are two examples (*64-66*).

1.2.4 Arginine-rich motif (ARM)

The Arginine-rich motif is only 10-20 amino acids long and contains approximately 4-8 Arg residues. The ARM is found in a variety of RNA-binding proteins, including the HIV and BIV Tat transcription factors, the bacteriophage antiterminator N proteins, viral

coat proteins and ribosomal proteins (*67-71*). In cells, proteins containing ARMs are involved mainly in the transcriptional regulation of gene expression. For example, Lambdoid N proteins, which contain arginine-rich N-terminal RNA-binding regions, act as transcriptional antitermination factors in gene expression (*72*). The *trans*-activation lentiviral regulatory protein Tat enhances transcriptional elongation of the promoter-proximal RNA polymerase II in mammalian immunodeficiency viruses (*73-74*). In contrast to the RRM and KH domains described above, ARMs share little sequence similarity besides the higher number of Arg residues and exhibit substantial diversity. The structures of the RNA targets can include hairpins, internal loops and bulges. ARMs use β-hairpin, α-helical and helix-turn-helix structures to bind RNA. The mechanisms mediating RNA recognition by ARMs have been investigated in depth in five ARM-RNA structures: BIV Tat-TAR (*75*), HIV-1 Rev-RRE (*76*), HIV-1 REV-RNA aptamer (*77*), P22-N-boxB (*78-79*), and HIV-1 TAR-argininamide (*80-81*). The diverse structures of ARMs enable them to act as essential building units to generate precise and specific RNA binders (Figure 1-4). The ARM usually binds to RNA in widened major grooves by mutual induce-fit mechanisms (*82*). Although the molecular mechanisms for major-groove widening are different in each complex, they all take advantage of irregular RNA elements, such as mismatches, loops and bulges, to achieve the same goal. For example, the major grooves in HIV-1 and HIV-2 TAR are widened by the buckled triple bases UAU and a projected unpaired uridine. The resulting local distortion makes a deep peptide-binding cavity in the widened major groove. The size of binding surface is further stretched by the phosphate backbone between G9 and U10. Interestingly, the local distortions of the helical structure have almost no effect on the overall base stacking of

RNA (*75, 83*). The major groove in HIV-1 RRE is widened by two successive G:A and G:G mismatches along the helical structure of RNA. A bulged uridine between the two mismatches partially unwinds the double helix and further widens the major groove of RRE (*76*).

The arginine side chain has the potential to form hydrogen bonds, weak polar, diverse ionic, and van der Waals interactions with RNA (*84-85*). In ARM-RNA complexes, the majority of the arginine residues in ARMs are found to contact loops, mismatches, and hairpins of RNA. For example, in the BIV Tat-TAR complex, the guanidinium groups of Arg70 and Arg73 form an extensive hydrogen bonding network with the major-groove nucleotides and the phosphodiester backbone of RNA in the central triple base region (*75*). In the Rev-aptamer complex, the aliphatic region of the side chain of Arg35 forms van der Waals interactions with the major-groove edge of RNA. Arg39 and Arg42 also make van der Waals contacts and a few hydrogen bonding interactions with different regions of the major-groove edge of RNA (*77*). Binding is not just mediated by arginine, the side chains of nonpolar and aromatic residues in ARMs are also involved in RNA recognition (*78, 86-87*). By contacting nonpolar pockets of the RNA or broad aromatic surfaces such as flipped bases or RNA loops, shape selectivity can be achieved.

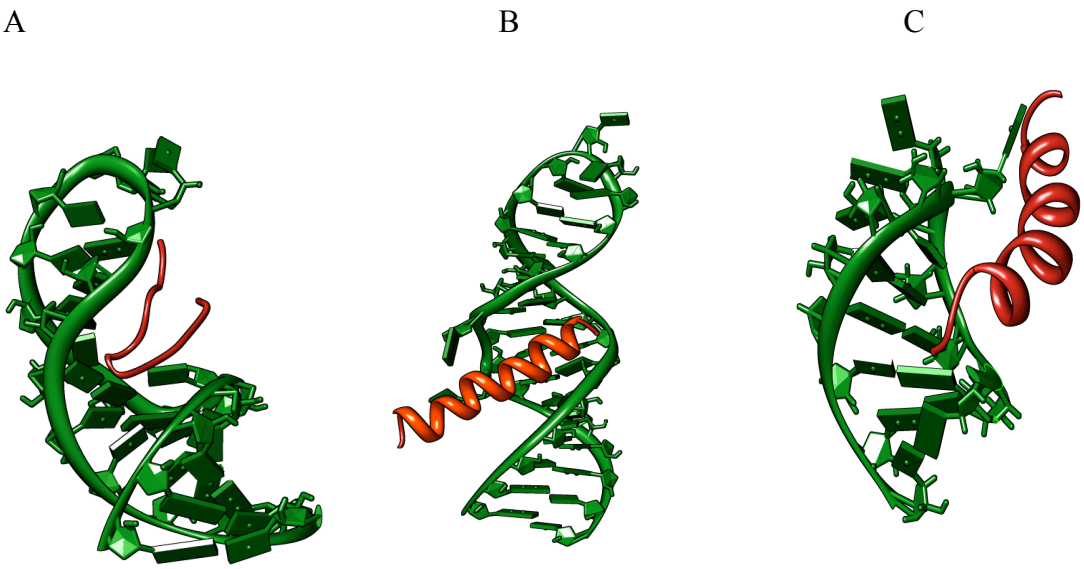


Figure 1-4. Structures of arginine-rich peptide–RNA complexes. A. BIV Tat–TAR RNA complex (PDB code 1MNB) B. HIV1 Rev–RRE RNA complex (PDB code 1ETF) C. P22 N–Box B RNA complex (PDB code 1A4T).

1.2.5 Other RNA-binding motifs

There are other RNA-binding motifs that have been identified. For example, the Arg-Gly-Gly box (RGG box) is comprised of approximately 20 amino acids. It is usually present as multiple copies and associated with other RNA binding motifs for RNA recognition, suggesting it may not recognize RNA sequences specifically (*88-91*). The motif is highly flexible due to the high percentage of Gly residues. Another important RNA-binding motif is the zinc-finger motif, which has been identified in many RNA-binding proteins, including the retroviral nucleocapsid protein, yeast RNA-binding protein, RNA polymerase III and TFIIIA protein in the ribosome (*92-96*). It folds into a compact zinc-binding domain with a shape of a finger. This motif is characterized by an α-helix and an anti-parallel β-sheet in a tandem pattern coordinated by zinc. Zinc fingers are generally present in multiple copies for RNA binding. The uniquely extended binding

surface allows it to bind to both single-stranded and double-stranded RNA (*92-94, 97*). Zinc fingers are classified into three families based on the sequences of the residues that use to coordinate zinc, namely CCHH (*94*), CCCH (*98*) and CCHC (*99*).

The RNA-binding motifs shown here provide important guidelines to classify RNA-binding proteins and outline the molecular basis that guides RNA recognition. Generally, protein-RNA complexes are recognized by hydrogen bonding, stacking and electrostatic interactions.

1.3 Conclusions

Besides these individual contacts described above, RNA-binding proteins also take advantage of cooperative interactions to obtain evolved binding affinities. We have identified and evaluated the effects of the local cooperativity of three neighboring residues, Asn15, Asn16 and Glu19, which contributes to the high affinity of the N-terminal RRM of the U1A protein upon binding to the U1 snRNA hairpin II loop. The localized cooperative network was mapped by double-mutant cycles and investigated with phage display. In the U1A protein, Asn15, Asn16 and Glu19 are located in β-strand 1 and loop 1 and contact U2, U3 and G4 of the RNA loop with specific hydrogen binding interactions. To probe the potential energetic coupling among these residues, we constructed a series of U1A mutants containing single, double, and triple-point Ala substitutions at positions 15, 16 and 19. The effects of these mutants on the overall stability of the U1A-RNA complex were individually determined. Double mutant cycle analysis revealed that Glu19 is energetically coupled with Asn15 and Asn16 by 0.4 kcal/mol and 0.5 kcal/mol, respectively, whereas there is virtually no energetic coupling between Asn15 and Asn16. We show that the U1A-RNA complex is stabilized by 0.9

kcal/mol due to the additive effect of energetic couplings between Asn15-Glu19 and Asn16 and Glu19. Phage display with randomization of three target sites demonstrates Asn15, Asn16 and Glu19 are preferred residues by U1A, and their signature side chains are required for obtaining the high affinity of the U1A-RNA complex. Our analysis also suggests that a cluster of these residues forms a "hot spot" on the surface of U1A. This cluster could therefore be used to rationally design small molecules or peptide mimics as regulators of the U1A-RNA complex. Overall, our studies complement other investigations of cooperative networks in U1A-RNA and provide insight towards the understanding of RNA binding by RRMs. Details of this project are described in Chapter 3.

1.4 References

1. Chinnusamy, V., Gong, Z., and Zhu, J. K. (2008) Nuclear RNA export and its importance in abiotic stress responses of plants, *Curr Top Microbiol Immunol 326*, 235-255.

2. Svoboda, P., and Di Cara, A. (2006) Hairpin RNA: a secondary structure of primary importance, *Cell Mol Life Sci 63*, 901-908.

3. Laurencikiene, J., Kallman, A. M., Fong, N., Bentley, D. L., and Ohman, M. (2006) RNA editing and alternative splicing: the importance of co-transcriptional coordination, *EMBO Rep 7*, 303-307.

4. Golden, D. E., and Hajduk, S. L. (2006) The importance of RNA structure in RNA editing and a potential proofreading mechanism for correct guide RNA:pre-mRNA binary complex formation, *J Mol Biol 359*, 585-596.

5. Heidrich, N., and Brantl, S. (2003) Antisense-RNA mediated transcriptional attenuation: importance of a U-turn loop structure in the target RNA of plasmid pIP501 for efficient inhibition by the antisense RNA, *J Mol Biol 333*, 917-929.

6. Dreyfuss, G., Kim, V. N., and Kataoka, N. (2002) Messenger-RNA-binding proteins and the messages they carry, *Nat Rev Mol Cell Biol 3*, 195-205.

7. Felciano, R. M., Chen, R. O., and Altman, R. B. (1997) RNA secondary structure as a reusable interface to biological information resources, *Gene 190*, GC59-70.

8. Ule, J. (2008) Ribonucleoprotein complexes in neurologic diseases, *Curr Opin Neurobiol 18*, 516-523.

9. Oeffinger, M., Wei, K. E., Rogers, R., DeGrasse, J. A., Chait, B. T., Aitchison, J. D., and Rout, M. P. (2007) Comprehensive analysis of diverse ribonucleoprotein complexes, *Nat Methods 4*, 951-956.

10. Luque, D., Saugar, I., Rejas, M. T., Carrascosa, J. L., Rodriguez, J. F., and Caston, J. R. (2009) Infectious Bursal disease virus: ribonucleoprotein complexes of a double-stranded RNA virus, *J Mol Biol 386*, 891-901.

11. Matsushima, Y., Matsumura, K., and Kitagawa, Y. (1997) Zinc finger-like motif conserved in a family of RNA binding proteins, *Biosci Biotechnol Biochem 61*, 905-906.

12. Siomi, H., and Dreyfuss, G. (1997) RNA-binding proteins as regulators of gene expression, *Curr Opin Genet Dev 7*, 345-353.

13. Arnez, J. G., and Cavarelli, J. (1997) Structures of RNA-binding proteins, *Q Rev Biophys 30*, 195-240.

14. Zamore, P. D., Williamson, J. R., and Lehmann, R. (1997) The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins, *RNA 3*, 1421-1433.

15. Antson, A. A. (2000) Single-stranded-RNA binding proteins, *Curr Opin Struct Biol 10*, 87-94.

16. Clarke, P. A., and Mathews, M. B. (1995) Interactions between the double-stranded RNA binding motif and RNA: definition of the binding site for the interferon-induced protein kinase DAI (PKR) on adenovirus VA RNA, *RNA 1*, 7-20.

17. Schmedt, C., Green, S. R., Manche, L., Taylor, D. R., Ma, Y., and Mathews, M. B. (1995) Functional characterization of the RNA-binding domain and motif of the double-stranded RNA-dependent protein kinase DAI (PKR), *J Mol Biol 249*, 29-44.

18. Adinolfi, S., Bagni, C., Castiglione Morelli, M. A., Fraternali, F., Musco, G., and Pastore, A. (1999) Novel RNA-binding motif: the KH module, *Biopolymers 51*, 153-164.

19. Tian, B., Bevilacqua, P. C., Diegelman-Parente, A., and Mathews, M. B. (2004) The double-stranded-RNA-binding motif: interference and much more, *Nat Rev Mol Cell Biol 5*, 1013-1023.

20. Burd, C. G., and Dreyfuss, G. (1994) Conserved structures and diversity of functions of RNA-binding proteins, *Science 265*, 615-621.

21. Messias, A. C., and Sattler, M. (2004) Structural basis of single-stranded RNA recognition, *Acc Chem Res 37*, 279-287.

22. Varani, G., and Nagai, K. (1998) RNA recognition by RNP proteins during RNA processing, *Annu Rev Biophys Biomol Struct 27*, 407-445.

23. Hall, K. B. (1994) Interaction of RNA hairpins with the human U1A N-terminal RNA binding domain, *Biochemistry 33*, 10076-10088.

24. Finn, R. D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, E. L., and Bateman, A. (2006) Pfam: clans, web tools and services, *Nucleic Acids Res 34*, D247-251.

25. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994) Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin, *Nature 372*, 432-438.

26. Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999) Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein, *Nature 398*, 579-585.

27. Law, M. J., Chambers, E. J., Katsamba, P. S., Haworth, I. S., and Laird-Offringa, I. A. (2005) Kinetic analysis of the role of the tyrosine 13, phenylalanine 56 and glutamine 54 network in the U1A/U1 hairpin II interaction, *Nucleic Acids Res 33*, 2917-2928.

28. Stark, H., Dube, P., Luhrmann, R., and Kastner, B. (2001) Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle, *Nature 409*, 539-542.

29. Wang, X., and Tanaka Hall, T. M. (2001) Structural basis for recognition of AU-rich element RNA by the HuD protein, *Nat Struct Biol 8*, 141-145.

30. Auweter, S. D., Oberstrass, F. C., and Allain, F. H. (2006) Sequence-specific binding of single-stranded RNA: is there a code for recognition?, *Nucleic Acids Res 34*, 4943-4959.

31. Maris, C., Dominguez, C., and Allain, F. H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression, *FEBS J 272*, 2118-2131.

32. Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A., and Sattler, M. (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP, *Mol Cell 11*, 965-976.

33. Kielkopf, C. L., Rodionova, N. A., Green, M. R., and Burley, S. K. (2001) A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer, *Cell 106*, 595-605.

34. Siomi, H., Matunis, M. J., Michael, W. M., and Dreyfuss, G. (1993) The pre-mRNA binding K protein contains a novel evolutionarily conserved motif, *Nucleic Acids Res 21*, 1193-1198.

35. Backe, P. H., Messias, A. C., Ravelli, R. B., Sattler, M., and Cusack, S. (2005) X-ray crystallographic and NMR studies of the third KH domain of hnRNP K in complex with single-stranded nucleic acids, *Structure 13*, 1055-1067.

36. Braddock, D. T., Baber, J. L., Levens, D., and Clore, G. M. (2002) Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: solution structure of a complex between hnRNP K KH3 and single-stranded DNA, *EMBO J 21*, 3476-3485.

37. Jensen, K. B., Musunuru, K., Lewis, H. A., Burley, S. K., and Darnell, R. B. (2000) The tetranucleotide UCAY directs the specific recognition of RNA by the Nova K-homology 3 domain, *Proc Natl Acad Sci U S A 97*, 5740-5745.

38. De Boulle, K., Verkerk, A. J., Reyniers, E., Vits, L., Hendrickx, J., Van Roy, B., Van den Bos, F., de Graaff, E., Oostra, B. A., and Willems, P. J. (1993) A point mutation in the FMR-1 gene associated with fragile X mental retardation, *Nat Genet 3*, 31-35.

39. Grishin, N. V. (2001) KH domain: one motif, two folds, *Nucleic Acids Res 29*, 638-643.

40. Lewis, H. A., Musunuru, K., Jensen, K. B., Edo, C., Chen, H., Darnell, R. B., and Burley, S. K. (2000) Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome, *Cell 100*, 323-332.

41. Lunde, B. M., Moore, C., and Varani, G. (2007) RNA-binding proteins: modular design for efficient function, *Nat Rev Mol Cell Biol 8*, 479-490.

42. Ostareck-Lederer, A., Ostareck, D. H., and Hentze, M. W. (1998) Cytoplasmic regulatory functions of the KH-domain proteins hnRNPs K and E1/E2, *Trends Biochem Sci 23*, 409-411.

43. Baber, J. L., Libutti, D., Levens, D., and Tjandra, N. (1999) High precision solution structure of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K, a c-myc transcription factor, *J Mol Biol 289*, 949-962.

44. Lewis, H. A., Chen, H., Edo, C., Buckanovich, R. J., Yang, Y. Y., Musunuru, K., Zhong, R., Darnell, R. B., and Burley, S. K. (1999) Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding domains, *Structure 7*, 191-203.

45. St Johnston, D., Brown, N. H., Gall, J. G., and Jantsch, M. (1992) A conserved double-stranded RNA-binding domain, *Proc Natl Acad Sci U S A 89*, 10979-10983.

46. McCormack, S. J., Thomis, D. C., and Samuel, C. E. (1992) Mechanism of interferon action: identification of a RNA binding domain within the N-terminal region of the human RNA-dependent P1/eIF-2 alpha protein kinase, *Virology 188*, 47-56.

47. Green, S. R., and Mathews, M. B. (1992) Two RNA-binding motifs in the double-stranded RNA-activated protein kinase, DAI, *Genes Dev 6*, 2478-2490.

48. Coggill, P., Finn, R. D., and Bateman, A. (2008) Identifying protein domains with the Pfam database, *Curr Protoc Bioinformatics Chapter 2*, Unit 2 5.

49. Bycroft, M., Grunert, S., Murzin, A. G., Proctor, M., and St Johnston, D. (1995) NMR solution structure of a dsRNA binding domain from Drosophila staufen protein reveals homology to the N-terminal domain of ribosomal protein S5, *EMBO J 14*, 3563-3571.

50. Kharrat, A., Macias, M. J., Gibson, T. J., Nilges, M., and Pastore, A. (1995) Structure of the dsRNA binding domain of E. coli RNase III, *EMBO J 14*, 3572-3584.

51. Nanduri, S., Carpick, B. W., Yang, Y., Williams, B. R., and Qin, J. (1998) Structure of the double-stranded RNA-binding domain of the protein kinase PKR reveals the molecular basis of its dsRNA-mediated activation, *EMBO J 17*, 5458-5465.

52. Ryter, J. M., and Schultz, S. C. (1998) Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA, *EMBO J 17*, 7505-7513.

53. Nicholson, A. W. (1996) Structure, reactivity, and biology of double-stranded RNA, *Prog Nucleic Acid Res Mol Biol 52*, 1-65.

54. Lehmann, K. A., and Bass, B. L. (1999) The importance of internal loops within RNA substrates of ADAR1, *J Mol Biol 291*, 1-13.

55. Krovat, B. C., and Jantsch, M. F. (1996) Comparative mutational analysis of the double-stranded RNA binding domains of Xenopus laevis RNA-binding protein A, *J Biol Chem 271*, 28112-28119.

56. Xu, M., Wells, K. S., and Emeson, R. B. (2006) Substrate-dependent contribution of double-stranded RNA-binding motifs to ADAR2 function, *Mol Biol Cell 17*, 3211-3220.

57. Stefl, R., Xu, M., Skrisovska, L., Emeson, R. B., and Allain, F. H. (2006) Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs, *Structure 14*, 345-355.

58. Leulliot, N., Quevillon-Cheruel, S., Graille, M., van Tilbeurgh, H., Leeper, T. C., Godin, K. S., Edwards, T. E., Sigurdsson, S. T., Rozenkrants, N., Nagel, R. J., Ares, M., and Varani, G. (2004) A new alpha-helical extension promotes RNA binding by the dsRBD of Rnt1p RNAse III, *EMBO J 23*, 2468-2477.

59. Ramos, A., Grunert, S., Adams, J., Micklem, D. R., Proctor, M. R., Freund, S., Bycroft, M., St Johnston, D., and Varani, G. (2000) RNA recognition by a Staufen double-stranded RNA-binding domain, *EMBO J 19*, 997-1009.

60. Wu, H., Henras, A., Chanfreau, G., and Feigon, J. (2004) Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III, *Proc Natl Acad Sci U S A 101*, 8307-8312.

61. Brendel, C., Rehbein, M., Kreienkamp, H. J., Buck, F., Richter, D., and Kindler, S. (2004) Characterization of Staufen 1 ribonucleoprotein complexes, *Biochem J 384*, 239-246.

62. Mouland, A. J., Mercier, J., Luo, M., Bernier, L., DesGroseillers, L., and Cohen, E. A. (2000) The double-stranded RNA-binding protein Staufen is incorporated in human immunodeficiency virus type 1: evidence for a role in genomic RNA encapsidation, *J Virol 74*, 5441-5451.

63. Stephens, O. M., Haudenschild, B. L., and Beal, P. A. (2004) The binding selectivity of ADAR2's dsRBMs contributes to RNA-editing selectivity, *Chem Biol 11*, 1239-1250.

64. Wang, W., Riedel, K., Lynch, P., Chien, C. Y., Montelione, G. T., and Krug, R. M. (1999) RNA binding by the novel helical domain of the influenza virus NS1 protein requires its dimer structure and a small number of specific basic amino acids, *RNA 5*, 195-205.

65. Silverman, R. H. (2007) Viral encounters with 2',5'-oligoadenylate synthetase and RNase L during the interferon antiviral response, *J Virol 81*, 12720-12729.

66. Hartmann, R., Norby, P. L., Martensen, P. M., Jorgensen, P., James, M. C., Jacobsen, C., Moestrup, S. K., Clemens, M. J., and Justesen, J. (1998) Activation of 2'-5' oligoadenylate synthetase by single-stranded and double-stranded RNA aptamers, *J Biol Chem 273*, 3236-3246.

67. Chaloin, L., Smagulova, F., Hariton-Gazal, E., Briant, L., Loyter, A., and Devaux, C. (2007) Potent inhibition of HIV-1 replication by backbone cyclic peptides bearing the Rev arginine rich motif, *J Biomed Sci 14*, 565-584.

68. Annamalai, P., Apte, S., Wilkens, S., and Rao, A. L. (2005) Deletion of highly conserved arginine-rich RNA binding motif in cowpea chlorotic mottle virus capsid protein results in virion structural alterations and RNA packaging constraints, *J Virol 79*, 3277-3288.

69. Venter, P. A., Marshall, D., and Schneemann, A. (2009) Dual roles for an arginine-rich motif in specific genome recognition and localization of viral coat protein to RNA replication sites in flock house virus-infected cells, *J Virol 83*, 2872-2882.

70. Lazinski, D., Grzadzielska, E., and Das, A. (1989) Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif, *Cell 59*, 207-218.

71. Hemmerich, P., Bosbach, S., von Mikecz, A., and Krawinkel, U. (1997) Human ribosomal protein L7 binds RNA with an alpha-helical arginine-rich and lysine-rich domain, *Eur J Biochem 245*, 549-556.

72. Doelling, J. H., and Franklin, N. C. (1989) Effects of all single base substitutions in the loop of boxB on antitermination of transcription by bacteriophage lambda's N protein, *Nucleic Acids Res 17*, 5565-5577.

73. Kato, H., Sumimoto, H., Pognonec, P., Chen, C. H., Rosen, C. A., and Roeder, R. G. (1992) HIV-1 Tat acts as a processivity factor in vitro in conjunction with cellular elongation factors, *Genes Dev 6*, 655-666.

74. Feinberg, M. B., Baltimore, D., and Frankel, A. D. (1991) The role of Tat in the human immunodeficiency virus life cycle indicates a primary effect on transcriptional elongation, *Proc Natl Acad Sci U S A 88*, 4045-4049.

75. Puglisi, J. D., Chen, L., Blanchard, S., and Frankel, A. D. (1995) Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex, *Science 270*, 1200-1203.

76. Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D., and Williamson, J. R. (1996) Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex, *Science 273*, 1547-1551.

77. Ye, X., Gorin, A., Ellington, A. D., and Patel, D. J. (1996) Deep penetration of an alpha-helix into a widened RNA major groove in the HIV-1 rev peptide-RNA aptamer complex, *Nat Struct Biol 3*, 1026-1033.

78. Legault, P., Li, J., Mogridge, J., Kay, L. E., and Greenblatt, J. (1998) NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif, *Cell 93*, 289-299.

79. Cai, Z., Gorin, A., Frederick, R., Ye, X., Hu, W., Majumdar, A., Kettani, A., and Patel, D. J. (1998) Solution structure of P22 transcriptional antitermination N peptide-boxB RNA complex, *Nat Struct Biol 5*, 203-212.

80. Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., and Williamson, J. R. (1992) Conformation of the TAR RNA-arginine complex by NMR spectroscopy, *Science 257*, 76-80.

81. Aboul-ela, F., Karn, J., and Varani, G. (1995) The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein, *J Mol Biol 253*, 313-332.

82. Tan, R., Chen, L., Buettner, J. A., Hudson, D., and Frankel, A. D. (1993) RNA recognition by an isolated alpha helix, *Cell 73*, 1031-1040.

83. Ye, X., Kumar, R. A., and Patel, D. J. (1995) Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex, *Chem Biol 2*, 827-840.

84. Patel, D. J. (1999) Adaptive recognition in RNA complexes with peptides and protein modules, *Curr Opin Struct Biol 9*, 74-87.

85. Calnan, B., Tidor, B., Biancalana, S., Hudson, D., and Frankel, A. (1991) Arginine-mediated RNA recognition: the arginine fork, *Science 252*, 1167-1171.

86. Chen, L., and Frankel, A. D. (1995) A peptide interaction in the major groove of RNA resembles protein interactions in the minor groove of DNA, *Proc Natl Acad Sci U S A 92*, 5077-5081.

87. Su, L., Radek, J. T., Hallenga, K., Hermanto, P., Chan, G., Labeots, L. A., and Weiss, M. A. (1997) RNA recognition by a bent alpha-helix regulates transcriptional antitermination in phage lambda, *Biochemistry 36*, 12722-12732.

88. Kiledjian, M., and Dreyfuss, G. (1992) Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box, *EMBO J 11*, 2655-2664.

89. Corley, S. M., and Gready, J. E. (2008) Identification of the RGG Box Motif in Shadoo: RNA-Binding and Signaling Roles?, *Bioinform Biol Insights 2*, 383-400.

90. Zanotti, K. J., Lackey, P. E., Evans, G. L., and Mihailescu, M. R. (2006) Thermodynamics of the fragile X mental retardation protein RGG box interactions with G quartet forming RNA, *Biochemistry 45*, 8319-8330.

91. Corbin-Lickfett, K. A., Chen, I. H., Cocco, M. J., and Sandri-Goldin, R. M. (2009) The HSV-1 ICP27 RGG box specifically binds flexible, GC-rich sequences but not G-quartet structures, *Nucleic Acids Res*.

92. De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element, *Science 279*, 384-388.

93. Lu, D., Searles, M. A., and Klug, A. (2003) Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition, *Nature 426*, 96-100.

94. Searles, M. A., Lu, D., and Klug, A. (2000) The role of the central zinc fingers of transcription factor IIIA in binding to 5 S RNA, *J Mol Biol 301*, 47-60.

95. Suntharalingam, M., Alcazar-Roman, A. R., and Wente, S. R. (2004) Nuclear export of the yeast mRNA-binding protein Nab2 is linked to a direct interaction with Gfd1 and to Gle1 function, *J Biol Chem 279*, 35384-35391.

96. Schuster, C., Myslinski, E., Krol, A., and Carbon, P. (1995) Staf, a novel zinc finger protein that activates the RNA polymerase III promoter of the selenocysteine tRNA gene, *EMBO J 14*, 3777-3787.

97. Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., Chancellor, K. J., Wu, Z. R., and Summers, M. F. (2000) NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition, *J Mol Biol 301*, 491-511.

98. Lai, W. S., Carballo, E., Thorn, J. M., Kennington, E. A., and Blackshear, P. J. (2000) Interactions of CCCH zinc finger proteins with mRNA. Binding of tristetraprolin-related zinc finger proteins to Au-rich elements and destabilization of mRNA, *J Biol Chem 275*, 17827-17837.

99. D'Souza, V., and Summers, M. F. (2004) Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus, *Nature 431*, 586-590.

CHAPTER 2

MODULAR DESIGN OF RNA-BINDING PROTEINS

2.1 Introduction

Associated with complex tertiary structures of RNA, RNA-binding proteins regulate various aspects of cellular and viral regulatory processes in RNA metabolism, including RNA capping, pre-mRNA splicing, polyadenylation, mRNA editing and stability (*1-2*). Many RNA-binding motifs have been identified on the basis of sequence similarities and structural topologies (*3*), including the RRM, the KH domain, the arginine-rich motif, and zinc-finger motifs. In eukaryotic cells, proteins containing RNA-binding motifs often use conserved binding mechanisms for RNA recognition (*4-7*). This observation and the importance of RNA in biological processes have generated great interest in creating artificial sequence-specific RNA-binding proteins using conserved RNA binding motifs as scaffolds. This design process would provide us a better understanding of the functions of RNA-binding proteins and the molecular basis of RNA recognition, and most importantly, it could allow ultimately gene expression to be manipulated at the RNA level for therapeutic uses.

Although many high-resolution protein-RNA complexes have been solved by X-ray crystallography and NMR spectroscopy (*8-17*), much less progress was made towards the design of RNA-binding proteins. In the past two decades, structure-based rational design and combinatorial selection approaches have been applied for designing RNA-binding proteins with limited success.

2.2 Structure-based design

2.2.1 Engineering Pumilio repeats for single-stranded RNA recognition

Proteins containing the Pumilio-homology domains (PUM-HD) or Puf domains act as translational regulators of gene expression in a wide variety of eukaryotes (yeast, plants and mammals) by binding to specific RNA sequences in the 3'- untranslated regions of mRNA (*18*). Puf proteins are usually associated with other regulatory proteins to control diverse developmental events including germ-line stem cell development, anterior patterning and neuronal function (*18-20*). The PUM-HD comprises eight consecutive sequence repeats, called Puf or PUM repeats, each 36 amino acids long, flanked by two conserved Puf –related regions at the N- and C-terminus. Puf proteins recognize specific RNA sequences, called nanos response elements (NREs) located at the 3'-UTR of mRNA (*18, 21-23*). The crystal structures of the fly PUM-HD in complex with fly NRE revealed that eight Puf repeats line up to form an extended curved binding surface for RNA recognition (*24*). All eight repeats are required for efficient and specific RNA binding, and each RNA base is individually recognized by three conserved amino acids in the second helix in each repeat. This unique binding pattern makes it an excellent scaffold for developing linear single-stranded RNA-binding proteins with desired binding properties.

Cheong and Hall applied site-directed mutagenesis of Homo sapiens Pumilio 1 homology domain (HsPUM1-HD) to produce mutants of this protein with altered RNA binding specificity (*25*). They constructed seven mutants of HsPUM1-HD by mutating two of the three side chains of the conserved residues that contact the targeted single RNA base (Figure 2-1).

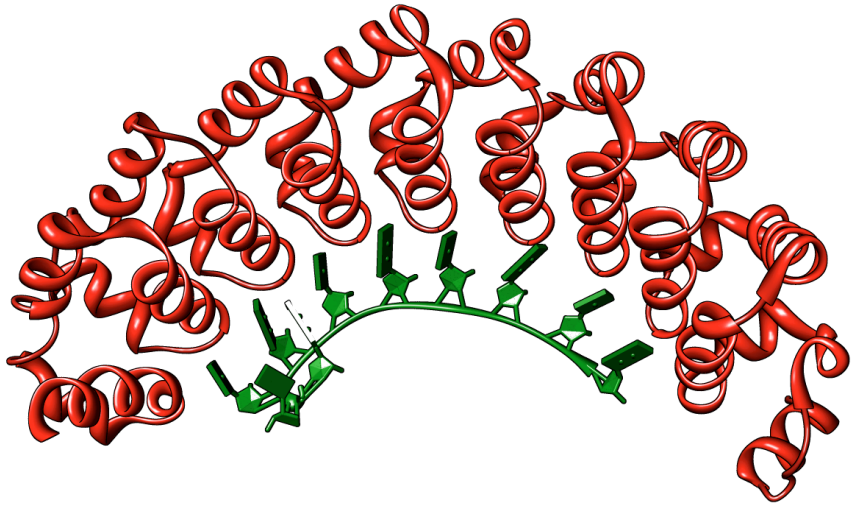


Figure 2-1. Schematic representation of sequence-specific NRE RNA recognition by HsPUM1-HD (PDB code 1M8Y).

They found that the two-point mutations of the protein are sufficient to switch the binding specificity from G to U, U to G, A to G and A to U. They did not mutate the site where ring stacking is present since it was thought to be important for affinity, not specificity (*25*). In many cases, the mutant proteins bind to the modified RNA targets with comparable affinity as does the wild-type protein to the wild-type NRE. In other cases, the complexes formed between the mutant proteins to their RNA targets were 10 to 30-fold less stable than the wild-type complex. These observations suggest recognition of NREs by Puf proteins is more complex and involves more residues than just the two residues involved in specific base contacts. Recognition of the backbone of the protein by water-mediated intermolecular hydrogen bonds and stacking interactions formed by alternating aromatic and basic side chains with the RNA bases also contribute to the affinity of the protein for RNA. A better understanding of these interactions and other factors will be necessary to generate high affinity single-stranded RNA-binding proteins by using the HsPUM-HD as a scaffold. Undoubtedly, these artificial RNA-binding

proteins will improve our understanding of translational regulation by 3'-UTR regulators and functions of other associated factor during cell development.

2.2.2 Design of a zinc finger that recognizes HIV-1 RRE

A structure based design approach has been used to generate an RNA-binding zinc finger that specifically recognizes the HIV-1 Rev response element (RRE) (*26*). Zinc fingers are commonly considered to be DNA-binding proteins, and they recognize DNA primarily by direct contacts with base pairs in the major groove (*5, 27-28*). Some residues in the α-helix (recognition code) of zinc fingers have been identified to be critical for DNA recognition (*29*). Several DNA-binding zinc fingers have been designed by selective mutations of these residues (*30-32*). Zinc fingers can also bind to RNA as shown by the RNA-binding domains in the transcription factor TFIIIA (*33*) and TIS11d (*34*). Other zinc fingers have been shown to bind to DNA-RNA hybrids (*35*).

McColl *et al.* designed a hybrid zinc-finger-Rev (ZF-Rev) peptide that mimics a general zinc finger and binds specifically to HIV-1 RRE (*26*). In the native complex of HIV-1 Rev-RRE, HIV-1 Rev utilizes an arginine-rich α-helix to recognize the RRE in the major groove. This arginine-rich α-helix is structurally unstable when it is isolated from the full-length protein HIV-1 Rev, and its affinity to the RRE is proportional to its helical content (*36*). It has been shown that this isolated arginine-rich peptide can be partially stabilized by extra peptide sequences added to the N terminus (*36*). A zinc finger peptide is an excellent candidate for locking the Rev peptide into a native helical state, which in turn increases its affinity to the RRE, since zinc finger peptides can fold into stable structures in a metal-dependant manner (*37-38*). In this study, the N-terminal portion (KPFQ**C**RI**C**MRNFS) of a single zinc finger domain of Zif268 (DNA-binding protein)

was engineered into an arginine-rich α-helix of Rev (TRQARRN**R**RRR**R**RR). Two non-essential Arg residues highlighted in the sequence were substituted by two His residues in the Rev peptide in order to construct the CCHH arrangement of the novel zinc finger. Cys residues in the zinc finger sequence required for zinc coordination and other critical residues required for RRE recognition in the Rev peptide were retained (*26*). The ZF-Rev hybrid peptide showed metal-binding capability and structural characteristics of standard zinc fingers by cobalt absorption and CD experiments. As expected, the novel hybrid zinc finger peptide exhibits the zinc-dependent binding to RRE, that is comparable to that of a stabilized isolated α-helical peptide (*36*). These findings suggest the designed hybrid zinc finger may bind to RRE with a stabilized α-helix similar to the wild-type Rev. Interestingly, the novel ZF-Rev peptide also showed binding activity in HeLa cells (*36*). A series of fusion proteins of the bacteriophage λ N protein with ZF-Rev fused to the N-terminus bound to RRE IIB in bacteria, suggesting the binding mode could be identical to that observed in mammalian cells.

This work demonstrates that the zinc finger motif can provide an excellent scaffold to design a monomeric sequence-specific RNA binding helix for HIV-1 RRE. In zinc finger-DNA complexes, multiple copies of zinc fingers are required for sequence-specific DNA recognition, because residues from adjacent fingers and interdomain linkers contact DNA bases in the major groove of DNA and help maintain the optimal orientation of α helix binding (*29*). In contrast to rigid DNA structures (A-form), RNA can form a wide variety of tertiary structures that are usually dynamic. Thus, RNA can potentially provide many defined binding sites, such as the deepened major groove in the RRE, for designed RNA binders with smaller sizes. The affinity and specificity of these binders could be

optimized by additional contacts with RNA, either in conjunction with the zinc-finger scaffold or perhaps within a scaffold customized by combinatorial selection (*26*).

2.2.3 Design of a tethered RNA-binding peptide

An individual RNA-binding motif usually has a limited ability to recognize RNA. The affinity and specificity of complex formation can be increased substantially by tethering RNA-binding domains together. Some examples are tandem zinc finger motifs (*39*), RNA recognition motifs (*40*) and the Pumilio protein (*23*). The various orientations of these tethered RNA-binding motifs expand the overall RNA-binding interface and thus, enable higher affinity and specificity of binding. There has been significant progress in designing RNA-binding peptides with multiple RNA-binding motifs linked together. For example, a novel sequence-specific RNA binding peptide, a dimeric covalently linked hairpin peptide, has been designed to bind to two adjacent TAR RNA sites (*41*). The bovine immunodeficiency virus (BIV) Tat peptide is unstructured, but forms a restricted hairpin when it binds to the major groove of BIV TAR. The NMR structure of the complex and mutational studies have identified eight amino acids within the 14-resiudes long arginine-rich motif of BIV Tat and two discontinuous G:C base pairs of BIV TAR that are important for sequence-specific RNA recognition. An additional G and a bulged U in the TAR sequence are also involved in RNA binding (*42-43*). In order to design a new RNA binder with higher affinity and specificity than its natural counterpart, two slightly modified native BIV Tat peptides (RGTRGKGRRI) were covalently linked together to recognize two " half-site" of BIV TAR.

The dimeric arrangement of two RNA-binding motifs can generate a high local concentration of the second motif when the first one is bound, resulting in cooperative

binding. The designed peptide binds 10-fold higher than the corresponding monomer when the two TAR sequences were placed directly adjacent to each other in a head-to-head dimeric arrangement (*41*). In this study, the designed tethered arginine-rich motifs were not able to achieve the expected binding affinity enhancement shown by tethered DNA-binding motifs (*31, 44*), suggesting the dimeric hairpin peptide conformation could be disordered and may need to be stabilized prior to binding. The stabilized α helical structure has been shown to increase the affinity of the RNA-binding domain of Rev for RNA, providing that the rigid structure of BIV peptide accommodates well into the RNA binding site (*36*). Additionally, the constructed TAR binding sites provide poor major groove accessibility because they only have a single bulge at the designated position. The results suggest that an arginine-rich BIV Tat peptide could be developed as a probe to investigate the accessibility of major grooves *in vivo*. Interestingly, asymmetric bulges and loops with extra nucleotides on the 5' side of the TAR binding site were found to create much more opened major grooves than when the extra nucleotides are introduced on the 3' side to TAR sites (*45*). These facts, together with optimization of the individual motifs and linkages, could be used to generate tethered RNA-binding motifs that bind RNA with high affinity and specificity.

2.2.4 Design of RNA-binding α helices with picomolar affinity

Another promising method for efficiently enhancing the stability and specificity of RNA- peptide complexes is to mix various critical residues from different binding motifs together to build a new RNA-binding peptide. A striking example is the structure-based design of a hybrid arginine-rich RNA-binding peptides that can bind to a hairpin stem with picomolar affinity (*46*).

A characteristic induced α helical structure ($P22_{N21}$) derived from bacteriophage binds to its target RNA, a P22boxB RNA hairpin, with ~200 pM affinity (*47-48*). This complex plays important roles in transcription antitermination of bacteriophage. Its counterpart antitermination complex in λ phage (λ_{N22} - λboxB) shows 100-fold less in affinity. The NMR solution structures of P22 and λ antitermination complexes with RNA show a similar binding pattern controlling both RNA-binding complexes (*49*). They both have arginine-rich peptides at their N-termini, which adopt bended α helical structures and insert into the major groove of the cognate RNA stem hairpins to make contacts with RNA base pairs (*50-54*). The pentaloops of P22boxB and λboxB hairpins are stabilized by a GNRA-fold in which one base is flipped out of the loops. Seven out eleven N-terminal residues of $P22_{N21}$ and λ_{N22} peptides are highly conserved and make direct contacts with the stem hairpins of cognate boxB RNA. In order to enhance the moderate stability observed for the λ_{N22}-λboxB complex, the peptide sequence information from the $P22_{N21}$ was extracted and applied to the sequence of λ_{N22}. In this study, the conserved residues were left intact to maintain the unique loop-binding surface, and only three non-conserved residues (Met, Asp and Gln) that contact with λboxB hairpin stem in λ_{N22} were substituted by reciprocal residues found in $P22_{N21}$. Arg at position 8 was not altered although there is a His at the similar position in $P22_{N21}$ (Figure 3-4). The binding affinities of mutated λ_{N22} peptides for λboxB are dramatically increased upon substitution of two or three of the residues that contact RNA. The resultant binding affinity is comparable to that of the counterpart $P22_{N21}$–P22boxB complex (*46*). Interestingly, the affinities are salt dependant. A change from 150 mM to 70 mM in salt concentrations increases the affinities by at least 10-fold depending on the sequences of the peptides.

The salt dependence data suggest these substituted residues contribute differently to the electrostatic interactions of the λ_{N22} - λboxB complex than they do to the $P22_{N21}$– P22boxB complex. Additionally, the substituted peptides show less helical character than that of the wild-type λ_{N22} in the absence of λboxB. The results of this study provide important implications for designing high affinity peptide binders to control cellular activities at the RNA level.

Although structure-based design has been proven to be successful in many cases, this approach requires a complete understanding of RNA recognition by proteins that has been limited by the handful of X-ray and NMR structures of protein-RNA complexes. Currently, studies on structure-based design focus on mainly small RNA-binding motifs such as zinc fingers and arginine-rich motifs in α helical or β hairpin conformations. The feasibility of designing proteins to recognize RNA tertiary structures using other RNA-binding motifs has yet explored.

2.3 Combinatorial functional selection

2.3.1 Selection of RNA-binding zinc fingers by phage display *in vitro*

Genetic selection of RNA-binding proteins from a combinatorial library provides much more promise for designing novel RNA-binding proteins, because little knowledge of the interactions is required in advance. The best known example is the selection of sequence-specific RNA binding proteins from a zinc finger phage library (*55*).

Zinc finger domains have been utilized to design sequence-specific DNA binders by phage display of combinatorial libraries and certain recognition codes have been elucidated (*30, 56*). Assuming these codes could be applicable to RNA recognition, Frisen and Darby presented an excellent example of selection of zinc finger proteins with

specific RNA binding properties (*55*). The scaffold they used was a tandem pair of zinc finger 4 isolated from the transcription factor III A (TFIIIA) in the ribosome. The single zinc finger is classified as CCHH (C_2H_2). TFIIIA has nine zinc fingers linked together to recognize both DNA and RNA. The two zinc fingers (4 and 6) in the middle are known to bind to 5S rRNA, although the structure of the complex is not known. Genetic studies have shown that the pattern of RNA recognition in the complex may be similar to that observed for DNA-binding fingers, in which a few critical residues in α-helix lock RNA in position by specific hydrogen binding interactions. In order to select residues that specifically contact 5S rRNA and HIV-1 Rev response element (RRE) IIB hairpin, they randomized eight positions between +1 and +10 in each α helix and two positions in the loop connecting the α helix and β strands in each zinc finger. They further randomized the linker between the two zinc fingers that allows flexibility of the orientation of each zinc finger (*57*). Residues on the β strands of each zinc finger were retained since they contribute to the overall stability of RNA binding and are unlikely to direct specific recognition of RNA as shown in the complex with DNA. By phage display *in vitro*, they identified at least two new zinc finger proteins that recognized their RNA targets. The binding affinities for both cases are comparable to that of the wild-type protein partners (*58*). Interestingly, the selected zinc fingers for HIV-1 RRE have no structural similarity with its native binding protein HIV-1 Rev. Among those zinc fingers bound to RRE IIB, the first zinc finger is rich in acidic and polar residues and the second zinc finger contains mainly basic residues such as Lys and Arg. It has been shown that the second zinc finger by itself binds to RRE IIB with similar affinity as two zinc fingers linked together. In the complex of HIV-1 Rev bound to RRE IIB, Rev uses an arginine-rich α helical domain to

recognize RRE IIB in the distorted major groove of RNA (*59*). The findings shown in this study suggest the zinc finger motif may mimic this binding pattern in the new complex. These selected arginine residues and other basic residues such as Lys may help orientate the RNA in the correct position, which in turn facilitates the formation of other contacts in the surrounding region. Using a larger RRE sequence (234 nucleotides) in the selections, phage display has recovered zinc finger binders with stronger affinity than the wild type counterpart, suggesting more contacts have been made by the extended RNA target (*60*).

These findings highlight the important role of α helices in determining the specificity of zinc finger binding. In combination with phage display, the zinc finger motif can be engineered to recognize different RNAs other than the natural targets. In this study, the RNA targets all contain widened major grooves by local distortions to accommodate α helical domains. It is still questionable if the combinatorial approach could be used for selection of zinc finger binding proteins for regular RNA duplexes or single-stranded RNA stretches, since they have no such binding sites. Other RNA-binding motifs such as the RRM, the KH domain, and the double-stranded RNA-binding motif could be more suitable for these RNA targets.

2.3.2 Selection of HIV TAR-binding peptides by retroviral replication *in vivo*

An arginine-rich (ARM) sequence (RKKRRQRRR) in HIV Tat requires cyclin T1 as a cofactor to form a stable, high-affinity RNA complex with a 3-nucleotide bulge in TAR (*61-63*). The Tat proteins in bovine immunodeficiency virus (BIV) and Jembrana disease virus (JDV) can bind to their cognate TAR sites independently despite the structural similarity of the TAR sites (*42, 64-66*). In these complexes, arginine-rich sequences

adopt a β-hairpin structure and recognize TAR sites specifically with side chains of non-arginine residues. The BIV Tat binds to HIV TAR with low affinity and cannot utilize cyclin T1 as a cofactor to increase binding and stability (*67*). In this study, a combinatorial peptide library in the context of an HIV-1 provirus was generated based on the BIV Tat ARM and a viral replication system was used to select for HIV TAR binders (*68*).

It has been pointed out that provirus containing the BIV Tar ARM and HIV TAR has no ability for replication due to the weak affinity, whereas those containing JDV Tar ARM and HIV TAR replicate efficiently (*69*). Due to these observations, the BIV Tat β-hairpin core sequence was intact in the library, while the first five N-terminal residues (SGPRP) that differ from JDV Tat ARM were randomized. The randomized residues were restricted to 12 types of residues by the designed degenerate codon XXY (X represents a mixture of A:G:C at a ratio 3:2.5:3, Y is a mixture of G:T at a ratio 2.5:3). This degenerate codon only allows for the expression of hydrophilic or charged residues as well as Ala, Pro and Gly. These residues are known for RNA recognition and provide conformational flexibility.

As expected, the peptides isolated from the combinatorial library can bind to HIV TAR with a pattern similar to that of the JDV Tat ARM. They all contain at least two basic residues in the randomized positions and require cyclin T1 as a cofactor for HIV TAR recognition (*68*). The identified peptide sequences represent a novel category of HIV TAR binding peptides and could be applied as inhibitors for therapeutic purposes. The rates of viral replication have been shown to correlate well with RNA-binding affinities of Tat-TAR interactions (*69*), suggesting it may be possible to incorporate other

combinatorial peptide libraries in the designated insertion site to identify high affinity TAR-binding peptides. Identification of peptide binders to RNAs other than TAR may also be possible by engineering TAR-substituted viruses, providing another reporter system correlated well with the targeted protein-RNA interaction.

2.3.3 Other genetic methods for selection of RNA-binding proteins

The size of phage display library is limited to complexities up to 10^8 by the transformation efficiency of *E. coli* cells. A larger library size may be beneficial for circumstances with more randomized positions. Recently, a much larger library (10^9-10^{11}) has been generated using a totally cell-free system, called *in vitro* compartmentalization (IVC) (*70*). In this study, authors took advantage of high affinity and specificity of the complex of zinc-finger motifs bound to DNA for the directed evolution of the human U1A protein. They fused the U1A protein library with randomization at position 13 and 56 to a six tandem zinc fingers isolated from Zif268 (1-3) and TFIIIA (1-3) at the N-terminus. The direct linkage between the genotype and phenotype was made via the binding complex upstream of the T7 promoter in the DNA template that drives protein expression. The expressed chimeric protein library can be released from the "artificial cells" after breaking the emulsions. These premises allow them to select the wild type and U1A mutants from the library with the native RNA target as the bait. The method could be adapted to select other customized RNA-binding proteins considering its simplicity and efficiency.

Ribosome display (*71*) and mRNA display (*72*) can generate even bigger complexity (10^{12}-10^{13}) *in vitro*. The applications of these two methods for selection of RNA-binding proteins have not yet been fully explored (*73-74*).

2.4 Future prospects

Many examples have shown that RNA-binding motifs provide excellent scaffolds to generate RNA-binding peptides and proteins with novel and interesting binding properties. These proteins alone or associated with other cellular proteins have many potential applications in terms of gene regulation. The efforts for designing novel RNA-binding proteins are currently limited to RNA structures containing stem-loops, internal loops and bulges. The use of more complex structures of RNA that are present in gene expression machinery such as those in the spliceosome and ribosome are still not practical. The ultimate challenge will be fulfilled with the development of new screening methods and complete understanding of protein-RNA interactions, along with more detailed characterization of thermodynamics and kinetics of protein-RNA complexes. Moreover, the development of novel peptide and protein scaffolds beyond the naturally occurring motifs for generating more diverse RNA binding surfaces is also important. Some substantial progresses have been made in the past few years in designing small-molecule scaffolds (*75-78*) and peptidomimetics (*79-80*) for RNA binding.

2.5 Conclusions

The conserved surface of the human U1A protein recognizes the U1 snRNA hairpin II loop during the initial spliceosomal assembly. In Chapter 4, we show the binding surface of U1A can be altered to recognize a non-cognate RNA. In the crystal structure of the U1A-RNA complex, Asn15, Asn16 and Glu19 are involved in the recognition of G4 of RNA and their important contributions to the high affinity of the U1A-RNA complex have been reported in Chapter 2. We assumed that these projected residues could be mutated to recognize RNA bases other than G. To test this hypothesis, we applied the

U1A phage library with randomization at positions 15, 16 and 19 constructed previously (see Chapter 2) for selecting binding proteins for RNA bearing G4 mutations. Two U1A mutants selected from the pooled U1A phage library were analyzed. Our data shows that a single residue substitution of Glu19 by Gln or His largely restores the binding affinity between U1A and the G4U RNA mutant. Asn15 and Asn16 were retained in our selections. Structural analysis and mutation studies support a new hydrogen bond formation on the disruptive interface of the U1A-RNA complex. A comparison of affinities of the wild-type U1A and the selected mutants for the wild type RNA reveals Glu at position 19 is crucial for the high specificity of U1A for G4 recognition. To our knowledge, this work is the first time that U1A has been redesigned/selected for non-natural RNA targets. Our findings suggest that U1A is a proof-of-concept for the design of RNA-binding proteins with desired binding properties.

2.6 References

1. Cusack, S. (1999) RNA-protein complexes, *Curr Opin Struct Biol 9*, 66-73.

2. Dreyfuss, G., Kim, V. N., and Kataoka, N. (2002) Messenger-RNA-binding proteins and the messages they carry, *Nat Rev Mol Cell Biol 3*, 195-205.

3. Burd, C. G., and Dreyfuss, G. (1994) Conserved structures and diversity of functions of RNA-binding proteins, *Science 265*, 615-621.

4. Maris, C., Dominguez, C., and Allain, F. H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression, *FEBS J 272*, 2118-2131.

5. Hall, T. M. (2005) Multiple modes of RNA recognition by zinc finger proteins, *Curr Opin Struct Biol 15*, 367-373.

6. Chang, K. Y., and Ramos, A. (2005) The double-stranded RNA-binding motif, a versatile macromolecular docking platform, *FEBS J 272*, 2109-2117.

7. Auweter, S. D., Oberstrass, F. C., and Allain, F. H. (2006) Sequence-specific binding of single-stranded RNA: is there a code for recognition?, *Nucleic Acids Res 34*, 4943-4959.

8. Hainzl, T., Huang, S., and Sauer-Eriksson, A. E. (2007) Interaction of signal-recognition particle 54 GTPase domain and signal-recognition particle RNA in the free signal-recognition particle, *Proc Natl Acad Sci U S A 104*, 14911-14916.

9. Walden, W. E., Selezneva, A. I., Dupuy, J., Volbeda, A., Fontecilla-Camps, J. C., Theil, E. C., and Volz, K. (2006) Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA, *Science 314*, 1903-1908.

10. Tishchenko, S., Nikonova, E., Nikulin, A., Nevskaya, N., Volchkov, S., Piendl, W., Garber, M., and Nikonov, S. (2006) Structure of the ribosomal protein L1-mRNA complex at 2.1 A resolution: common features of crystal packing of L1-RNA complexes, *Acta Crystallogr D Biol Crystallogr 62*, 1545-1554.

11. Gan, J., Tropea, J. E., Austin, B. P., Court, D. L., Waugh, D. S., and Ji, X. (2006) Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III, *Cell 124*, 355-366.

12. Nevskaya, N., Tishchenko, S., Volchkov, S., Kljashtorny, V., Nikonova, E., Nikonov, O., Nikulin, A., Kohrer, C., Piendl, W., Zimmermann, R., Stockley, P., Garber, M., and Nikonov, S. (2006) New insights into the interaction of ribosomal protein L1 with RNA, *J Mol Biol 355*, 747-759.

13. Hauenstein, S., Zhang, C. M., Hou, Y. M., and Perona, J. J. (2004) Shape-selective RNA recognition by cysteinyl-tRNA synthetase, *Nat Struct Mol Biol 11*, 1134-1141.

14. Beuth, B., Garcia-Mayoral, M. F., Taylor, I. A., and Ramos, A. (2007) Scaffold-independent analysis of RNA-protein interactions: the Nova-1 KH3-RNA complex, *J Am Chem Soc 129*, 10205-10210.

15. Auweter, S. D., Oberstrass, F. C., and Allain, F. H. (2007) Solving the structure of PTB in complex with pyrimidine tracts: an NMR study of protein-RNA complexes of weak affinities, *J Mol Biol 367*, 174-186.

16. Volpon, L., D'Orso, I., Young, C. R., Frasch, A. C., and Gehring, K. (2005) NMR structural study of TcUBP1, a single RRM domain protein from Trypanosoma cruzi: contribution of a beta hairpin to RNA binding, *Biochemistry 44*, 3708-3717.

17. Sidiqi, M., Wilce, J. A., Porter, C. J., Barker, A., Leedman, P. J., and Wilce, M. C. (2005) Formation of an alphaCP1-KH3 complex with UC-rich RNA, *Eur Biophys J 34*, 423-429.

18. Wickens, M., Bernstein, D. S., Kimble, J., and Parker, R. (2002) A PUF family portrait: 3'UTR regulation as a way of life, *Trends Genet 18*, 150-157.

19. Wharton, R. P., Sonoda, J., Lee, T., Patterson, M., and Murata, Y. (1998) The Pumilio RNA-binding domain is also a translational regulator, *Mol Cell 1*, 863-872.

20. Parisi, M., and Lin, H. (1999) The Drosophila pumilio gene encodes two functional protein isoforms that play multiple roles in germline development, gonadogenesis, oogenesis and embryogenesis, *Genetics 153*, 235-250.

21. Macdonald, P. M. (1992) The Drosophila pumilio gene: an unusually long transcription unit and an unusual protein, *Development 114*, 221-232.

22. Barker, D. D., Wang, C., Moore, J., Dickinson, L. K., and Lehmann, R. (1992) Pumilio is essential for function but not for distribution of the Drosophila abdominal determinant Nanos, *Genes Dev 6*, 2312-2326.

23. Zamore, P. D., Williamson, J. R., and Lehmann, R. (1997) The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins, *RNA 3*, 1421-1433.

24. Edwards, T. A., Pyle, S. E., Wharton, R. P., and Aggarwal, A. K. (2001) Structure of Pumilio reveals similarity between RNA and peptide binding motifs, *Cell 105*, 281-289.

25. Cheong, C. G., and Hall, T. M. (2006) Engineering RNA sequence specificity of Pumilio repeats, *Proc Natl Acad Sci U S A 103*, 13635-13639.

26. McColl, D. J., Honchell, C. D., and Frankel, A. D. (1999) Structure-based design of an RNA-binding zinc finger, *Proc Natl Acad Sci U S A 96*, 9521-9526.

27. Pavletich, N. P., and Pabo, C. O. (1993) Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers, *Science 261*, 1701-1707.

28. Nolte, R. T., Conlin, R. M., Harrison, S. C., and Brown, R. S. (1998) Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex, *Proc Natl Acad Sci U S A 95*, 2938-2943.

29. Wolfe, S. A., Nekludova, L., and Pabo, C. O. (2000) DNA recognition by Cys2His2 zinc finger proteins, *Annu Rev Biophys Biomol Struct 29*, 183-212.

30. Rebar, E. J., and Pabo, C. O. (1994) Zinc finger phage: affinity selection of fingers with new DNA-binding specificities, *Science 263*, 671-673.

31. Greisman, H. A., and Pabo, C. O. (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites, *Science 275*, 657-661.

32. Clemens, K. R., Wolf, V., McBryant, S. J., Zhang, P., Liao, X., Wright, P. E., and Gottesfeld, J. M. (1993) Molecular basis for specific recognition of both RNA and DNA by a zinc finger protein, *Science 260*, 530-533.

33. McBryant, S. J., Veldhoen, N., Gedulin, B., Leresche, A., Foster, M. P., Wright, P. E., Romaniuk, P. J., and Gottesfeld, J. M. (1995) Interaction of the RNA binding

fingers of Xenopus transcription factor IIIA with specific regions of 5 S ribosomal RNA, *J Mol Biol 248*, 44-57.

34. Hudson, B. P., Martinez-Yamout, M. A., Dyson, H. J., and Wright, P. E. (2004) Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d, *Nat Struct Mol Biol 11*, 257-264.

35. Shi, Y., and Berg, J. M. (1995) Specific DNA-RNA hybrid binding by zinc finger proteins, *Science 268*, 282-284.

36. Tan, R., Chen, L., Buettner, J. A., Hudson, D., and Frankel, A. D. (1993) RNA recognition by an isolated alpha helix, *Cell 73*, 1031-1040.

37. Parraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T., and Klevit, R. E. (1988) Zinc-dependent structure of a single-finger domain of yeast ADR1, *Science 241*, 1489-1492.

38. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., and Wright, P. E. (1989) Three-dimensional solution structure of a single zinc finger DNA-binding domain, *Science 245*, 635-637.

39. Lu, D., Searles, M. A., and Klug, A. (2003) Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition, *Nature 426*, 96-100.

40. Wang, X., and Tanaka Hall, T. M. (2001) Structural basis for recognition of AU-rich element RNA by the HuD protein, *Nat Struct Biol 8*, 141-145.

41. Campisi, D. M., Calabro, V., and Frankel, A. D. (2001) Structure-based design of a dimeric RNA-peptide complex, *EMBO J 20*, 178-186.

42. Chen, L., and Frankel, A. D. (1994) An RNA-binding peptide from bovine immunodeficiency virus Tat protein recognizes an unusual RNA structure, *Biochemistry 33*, 2708-2715.

43. Chen, L., and Frankel, A. D. (1995) A peptide interaction in the major groove of RNA resembles protein interactions in the minor groove of DNA, *Proc Natl Acad Sci U S A 92*, 5077-5081.

44. Robinson, C. R., and Sauer, R. T. (1998) Optimizing the stability of single-chain proteins by linker length and composition mutagenesis, *Proc Natl Acad Sci U S A 95*, 5929-5934.

45. Weeks, K. M., and Crothers, D. M. (1993) Major groove accessibility of RNA, *Science 261*, 1574-1577.

46. Austin, R. J., Xia, T., Ren, J., Takahashi, T. T., and Roberts, R. W. (2002) Designed arginine-rich RNA-binding peptides with picomolar affinity, *J Am Chem Soc 124*, 10966-10967.

47. Van Gilst, M. R., Rees, W. A., Das, A., and von Hippel, P. H. (1997) Complexes of N antitermination protein of phage lambda with specific and nonspecific RNA target sites on the nascent transcript, *Biochemistry 36*, 1514-1524.

48. Frankel, A. D. (2000) Fitting peptides into the RNA world, *Curr Opin Struct Biol 10*, 332-340.

49. Heus, H. A., and Pardi, A. (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops, *Science 253*, 191-194.

50. Scharpf, M., Sticht, H., Schweimer, K., Boehm, M., Hoffmann, S., and Rosch, P. (2000) Antitermination in bacteriophage lambda. The structure of the N36 peptide-boxB RNA complex, *Eur J Biochem 267*, 2397-2408.

51. Legault, P., Li, J., Mogridge, J., Kay, L. E., and Greenblatt, J. (1998) NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif, *Cell 93*, 289-299.

52. Weiss, M. A. (1998) RNA-mediated signaling in transcription, *Nat Struct Biol 5*, 329-333.

53. Patel, D. J. (1999) Adaptive recognition in RNA complexes with peptides and protein modules, *Curr Opin Struct Biol 9*, 74-87.

54. Weisberg, R. A., and Gottesman, M. E. (1999) Processive antitermination, *J Bacteriol 181*, 359-367.

55. Friesen, W. J., and Darby, M. K. (1998) Specific RNA binding proteins constructed from zinc fingers, *Nat Struct Biol 5*, 543-546.

56. Rebar, E. J., Greisman, H. A., and Pabo, C. O. (1996) Phage display methods for selecting zinc finger proteins with novel DNA-binding specificities, *Methods Enzymol 267*, 129-149.

57. Nakaseko, Y., Neuhaus, D., Klug, A., and Rhodes, D. (1992) Adjacent zinc-finger motifs in multiple zinc-finger peptides from SWI5 form structurally independent, flexibly linked domains, *J Mol Biol 228*, 619-636.

58. Mann, D. A., Mikaelian, I., Zemmel, R. W., Green, S. M., Lowe, A. D., Kimura, T., Singh, M., Butler, P. J., Gait, M. J., and Karn, J. (1994) A molecular rheostat. Co-operative rev binding to stem I of the rev-response element modulates human immunodeficiency virus type-1 late gene expression, *J Mol Biol 241*, 193-207.

59. Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D., and Williamson, J. R. (1996) Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex, *Science 273*, 1547-1551.

60. Friesen, W. J., and Darby, M. K. (2001) Specific RNA binding by a single C2H2 zinc finger, *J Biol Chem 276*, 1968-1973.

61. Bieniasz, P. D., Grdina, T. A., Bogerd, H. P., and Cullen, B. R. (1998) Recruitment of a protein complex containing Tat and cyclin T1 to TAR governs the species specificity of HIV-1 Tat, *EMBO J 17*, 7056-7065.

62. Chen, D., Fong, Y., and Zhou, Q. (1999) Specific interaction of Tat with the human but not rodent P-TEFb complex mediates the species-specific Tat activation of HIV-1 transcription, *Proc Natl Acad Sci U S A 96*, 2728-2733.

63. Ivanov, D., Kwak, Y. T., Nee, E., Guo, J., Garcia-Martinez, L. F., and Gaynor, R. B. (1999) Cyclin T1 domains involved in complex formation with Tat and TAR RNA are critical for tat-activation, *J Mol Biol 288*, 41-56.

64. Puglisi, J. D., Chen, L., Blanchard, S., and Frankel, A. D. (1995) Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex, *Science 270*, 1200-1203.

65. Smith, C. A., Calabro, V., and Frankel, A. D. (2000) An RNA-binding chameleon, *Mol Cell 6*, 1067-1076.

66. Ye, X., Kumar, R. A., and Patel, D. J. (1995) Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex, *Chem Biol 2*, 827-840.

67. Smith, C. A., Crotty, S., Harada, Y., and Frankel, A. D. (1998) Altering the context of an RNA bulge switches the binding specificities of two viral Tat proteins, *Biochemistry 37*, 10808-10814.

68. Xie, B., Calabro, V., Wainberg, M. A., and Frankel, A. D. (2004) Selection of TAR RNA-binding chameleon peptides by using a retroviral replication system, *J Virol 78*, 1456-1463.

69. Xie, B., Wainberg, M. A., and Frankel, A. D. (2003) Replication of human immunodeficiency viruses engineered with heterologous Tat-transactivation response element interactions, *J Virol 77*, 1984-1991.

70. Chen, Y., Mandic, J., and Varani, G. (2008) Cell-free selection of RNA-binding proteins using in vitro compartmentalization, *Nucleic Acids Res 36*, e128.

71. He, M., and Taussig, M. J. (2002) Ribosome display: cell-free protein display technology, *Brief Funct Genomic Proteomic 1*, 204-212.

72. Zhang, J. S., Duncan, E. L., Chang, A. C., and Reddel, R. R. (1998) Differential display of mRNA, *Mol Biotechnol 10*, 155-165.

73. Barrick, J. E., Takahashi, T. T., Ren, J., Xia, T., and Roberts, R. W. (2001) Large libraries reveal diverse solutions to an RNA recognition problem, *Proc Natl Acad Sci U S A 98*, 12374-12378.

74. Yan, S., Niu, R., Wang, Z., and Lin, X. (2007) In vitro selected peptides bind with thymidylate synthase mRNA and inhibit its translation, *Sci China C Life Sci 50*, 630-636.

75. Hermann, T. (2000) Strategies for the Design of Drugs Targeting RNA and RNA-Protein Complexes, *Angew Chem Int Ed Engl 39*, 1890-1904.

76. Sucheck, S. J., and Wong, C. H. (2000) RNA as a target for small molecules, *Curr Opin Chem Biol 4*, 678-686.

77. Zapp, M. L., Young, D. W., Kumar, A., Singh, R., Boykin, D. W., Wilson, W. D., and Green, M. R. (1997) Modulation of the Rev-RRE interaction by aromatic heterocyclic compounds, *Bioorg Med Chem 5*, 1149-1155.

78. Thomas, J. R., and Hergenrother, P. J. (2008) Targeting RNA with small molecules, *Chem Rev 108*, 1171-1224.

79. Leeper, T. C., Athanassiou, Z., Dias, R. L., Robinson, J. A., and Varani, G. (2005) TAR RNA recognition by a cyclic peptidomimetic of Tat protein, *Biochemistry 44*, 12362-12372.

80. Li, K., Fernandez-Saiz, M., Rigl, C. T., Kumar, A., Ragunathan, K. G., McConnaughie, A. W., Boykin, D. W., Schneider, H. J., and Wilson, W. D. (1997) Design and analysis of molecular motifs for specific recognition of RNA, *Bioorg Med Chem 5*, 1157-1172.

CHAPTER 3

IDENTIFICATION AND ANALYSIS OF LOCAL COOPERATIVE INTERACTIONS IN THE U1A-RNA COMPLEX

3.1 Introduction

The regulatory roles of RNA are highly dependent upon specific interactions with RNA binding proteins (*1-3*). In plants, animals and bacterial cells, most of these RNA-binding proteins have modular structures defined by different structural motifs. Single or multiple copies of these motifs allow specific recognition of RNA molecules diverse in structure and sequence (*4-8*). The RNA recognition motif (RRM), also referred to as the ribonucleoprotein domain (RNP), is one of the most abundant RNA-binding motifs. It recognizes single stranded RNA sequences and is involved in almost all steps of gene regulation, including, but not limited to, pre-mRNA splicing, transport, and stability (*9*). The RRM folds into a globular structure comprised of a four-stranded anti-parallel β-sheet supported on one face by two α-helices. The exposed β-sheet is the major binding surface for RNA recognition by most RRMs. RRM-RNA complexes are stabilized by conserved stacking interactions and large hydrogen bonding networks (*10*). RRMs are capable of discriminating RNA targets in cells (*11-12*). Through evolution, these RRMs developed to bind cognate RNAs with different affinities in order to execute their various functions. Therefore, understanding the molecular basis of RRM-RNA recognition is important for elucidating their individual roles in biological processes.

Residue cooperativity is considered a fundamental component in protein folding, enzyme catalysis and molecular recognition (*13-14*). In a binding interface, individual residues act cooperatively or are energetically coupled if the effect on complex stability

of modifying these residues simultaneously does not equal the sum of the effect on complex stability of modifying each residue individually (*15-17*). In this study, we quantitatively probed the cooperative interactions of three neighboring residues in the N-terminal RRM of the U1A protein upon binding to the U1 snRNA hairpin II loop. U1A is an excellent model system for studying RNA recognition by a single RRM.

The U1A protein is a key component of the U1 small nuclear ribonucleoprotein (U1 snRNP) in the spliceosome. It binds to a single-stranded RNA element AUUGCAC in the U1 snRNA hairpin II loop with very high affinity ($K_D \sim 10^{-11}$ M) (Figure 3-1) (*18-19*). U1A also binds to a similar sequence in two adjacent internal bulges in the 3'-untranslated regions of its own pre-mRNA. Although U1A contains two RRMs, the N-terminal RRM of U1A, containing the first 101 residues, confers full binding affinity to RNA (*10, 20*).

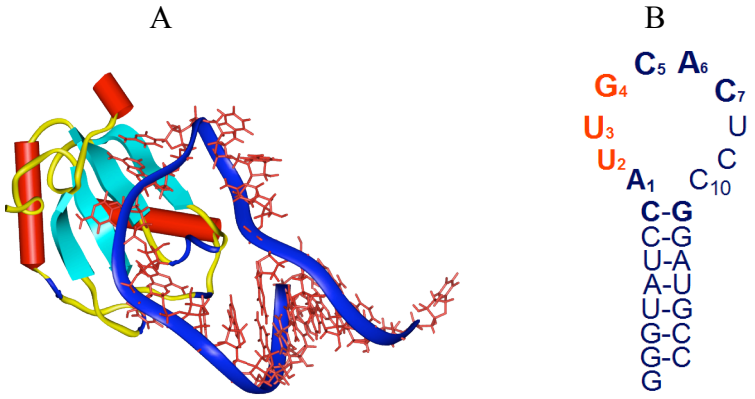


Figure 3-1. (A) N-terminal RRM of U1A in complex with U1 snRNA hairpin II (PDB entry 1URN). The U1 snRNA hairpin II is shown as the blue ribbon. (B) Secondary structure of U1 snRNA hairpin II. The recognition sequence is shown in boldface and U2, U3 and G4 for this study are highlighted in red.

Previous studies have revealed several long-range cooperative networks existing in the N terminal RRM of U1A that enhance binding to RNA and possibly increase the selectivity of RNA recognition (*21-23*). In one of these networks, Tyr13, a highly

conserved aromatic residue in β-strand 1, was found to be energetically coupled with Phe56, a highly conserved aromatic residue in β-strand 3. Moreover, both Tyr13 and Phe56 are coupled with most of the C-terminal residues of the U1A protein (*21*). The C-terminal region of the N-terminal RRM of U1A is comprised of β-strand 4 and α-helix C and forms an extensive hydrogen bonding network with the last three bases C5, A6 and C7 in the recognition element of RNA upon binding (*10*). In a second cooperative network, Tyr13 is energetically coupled with Gln54 and residues in loop 3 mediated through the G-C closing base pair and A1 of RNA (*22*). In a third cooperative network, Gly53 was found to be coupled with the TDS linker of U1A through loops 1 and 3 (*23*). Recently, a more complex cooperative network with the U1A-RNA complex has been reported from our group (*24*). All of these cooperative networks have focused largely on long-range energetic coupling of residues that describe the global rearrangement of U1A upon binding to RNA.

In this study, we have chosen to investigate local cooperativity in the U1A-RNA complex involving a cluster containing three neighboring residues (Asn15, Asn16 and Glu19). In the crystal structure of the complex, these residues make specific hydrogen bonding contacts with RNA and probably contribute significantly to the highly specific recognition of RNA by U1A (Figure 3-2) (*10*). Mutational studies have shown that these sites are important contributors to the high stability of the U1A-RNA complex (*25*). In order to the probe whether these residues act independently or in a cooperative manner, we exhaustively mutated Asn15, Asn16 and Glu19 to Ala. With the hypothesis that these mutations only remove local contacts without disturbing the entire binding interface and stability of the U1A-RNA complex, the alterations of the free binding energy upon

mutation represent the contributions from those deleted interactions. The energetic coupling of these residues was estimated by the method of double mutant cycles (*14*). We also used phage display (*26*) to investigate if the identities of Asn15, Asn16 and Glu19 are required for RNA binding.

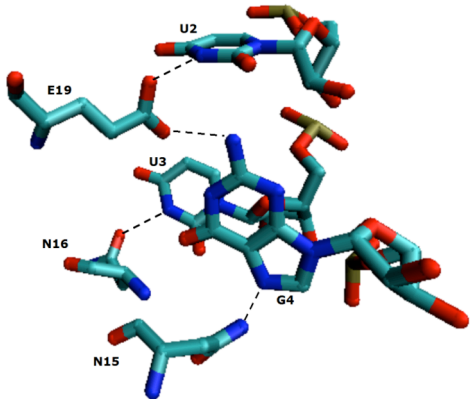


Figure 3-2. Structural representation of hydrogen bonding interactions between Asn 15, Asn 16 and Glu 19 of U1A and U2, U3 and G4 of U1 snRNA hairpin II . For clarity, only the side chains of the residues and splayed-out bases U3, U4 and G4 are shown. The hydrogen bonds investigated in this study are shown by dashed lines. The two nonspecific hydrogen bonds between G4 and the amide and carbonyl groups of the protein backbone in this region are not shown.

3.2 Results and discussion

We have investigated RNA recognition by the N-terminal RRM of U1A (residues 2-102), because it constitutes the portion of the U1A protein involved in binding the U1 snRNA hairpin II loop (*10*). In order to identify individual cooperative interactions within the locally cooperative network formed by Asn15, Asn16 and Glu19 and estimate their contributions to the overall stability of U1A-RNA, we constructed U1A mutants containing single, double and triple-point Ala substitutions at positions 15, 16 and 19. U1A mutants were expressed via a His_6 fusion in similar amounts in *E. coli* cells as the wild type U1A (see materials and methods), suggesting the stability of the proteins was

not dramatically affected by these mutations. Their affinities for RNA and binding free energies in comparison with that of the wild type U1A were determined (Figure 3-4) and are tabulated in Table 3-1.

3.2.1 Ala substitutions for Asn15, Asn16, and Glu19 in U1A

In the crystal structure of the U1A-RNA complex, Asn15 and Asn16 reside in β-strand 1 (Figure 2-4). The side chains of Asn15 and Asn16 make a single hydrogen bond with G4 and U3 of RNA, respectively (Figure 3-2). Glu19 is located in loop 1 connecting β-strand 1 and α-helix A. This loop contains residues 17-21 of U1A and Glu19 is placed in the middle, flanked by two residues on each side (Figure 3-3) (*10*). The positioning of Glu19 in the loop could be important for its optimal orientation for access to RNA. Glu19 forms two hydrogen bonds via its carboxyl group of the side chain with U2 and G4 of RNA (Figure 2-3) (*10*). Mutations in this region have been shown to greatly destabilize the complex (*25*). These residues do not appear to contribute to the tertiary structure of U1A and are exposed to solvent in the free protein. Thus, they are good candidates for investigating the contribution of cooperative interactions to RNA binding (*10, 25*).

```
1                   10                  20                  30                  40
M A V P E T R P N H T I Y I N N L N E K I K K D E L K K S L Y A I F S Q F G Q I
                    β1                              αA
                  50                  60                  70
L D I L V S R S L K M R G Q A F V I F K E V S S A T N A L R S M Q G F P F Y D
β1                          β3                  αB
80                  90                  100
K P M R I Q Y A K T D S D I I A K M K G T F V
    β3                  αC
```

Figure 3-3. Amino acid sequence and secondary structure of the N-terminal RRM of the U1A protein. Residue numbers are labeled above the sequence.

.

To investigate the individual contribution of Asn15, Asn16, and Glu19 to binding, we mutated each residue to Ala and measured the affinity of the resulting mutant protein for RNA. Due to their positions in the U1A structure, each mutation will eliminate the hydrogen bonding interactions involving that amino acid and may also alter the structure and stability of the free protein or complex. Thus, the observed destabilization and energetic coupling results from the combination of all of these effects on binding (*14*). The U1A protein and its mutants were expressed with His_6 tags. Under our binding conditions, the wild type His_6 fusion U1A binds to RNA with an affinity of 0.049 nM (Table 3-1), consistent with reported K_D values for the wild type complex without the tag (*19, 21-22*), suggesting the His_6 tag at the N-terminus does not interfere with the stability of the complex.

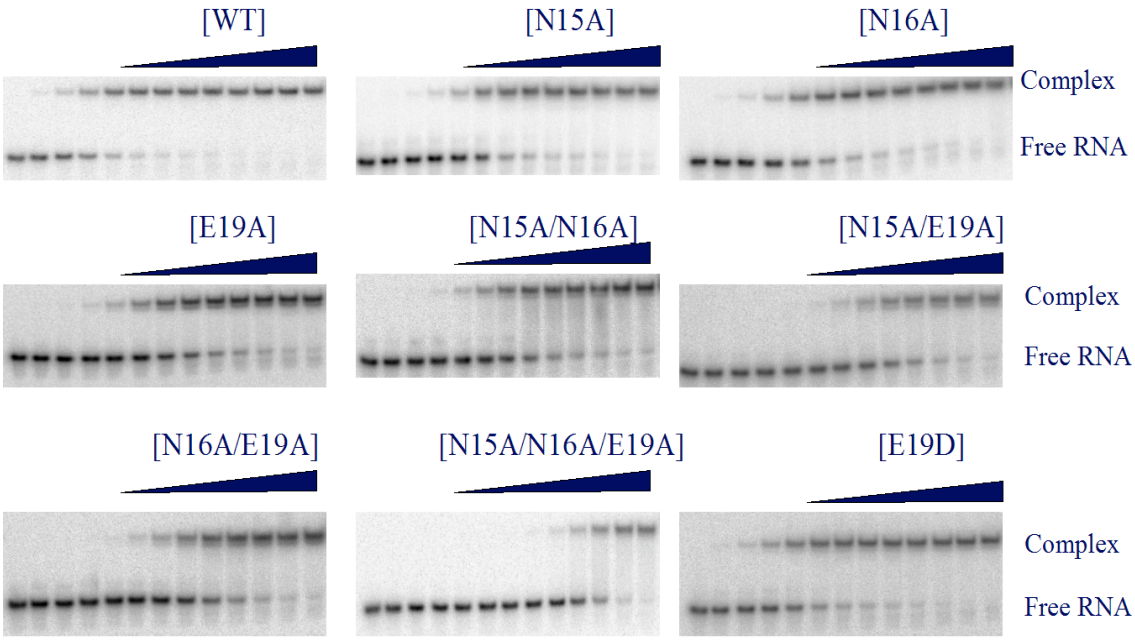


Figure 3-4. Representative electrophoretic mobility shift assays of U1A and its mutants in complex with the U1 snRNA hairpin II loop. Samples from a 4-fold serial dilution of the indicated protein were incubated with ^{32}P-labeled U1 snRNA hairpin II at room temperature for 30 min. The shifted bands were resolved by electrophoresis on an 8 % native polyacrylamide gel. Control reactions without added protein were loaded in the first lane on the left side of each gel.

As expected, all of the single Ala mutants formed complexes with RNA and were less stable than the wild type complex, suggesting the local arrangements of these residues have been evolved to recognize RNA. The calculated K_D for the complex formed with U1A mutant N15A was 0.5 nM, which is 10-fold higher than the K_D of the wild type complex. The corresponding energetic destabilization of the complex formed with U1A mutant N15A was 1.3 kcal/mol. Previous studies showed that a single substitution with Val at this position completely abolishes RNA binding (*25*). This could be due to a structural disturbance by the bulky side chain of Val and the different experimental conditions. A single substitution of Asn16 with Ala resulted in a 1.0 kcal/mol loss in the

binding free energy, slightly less than that resulting from N15A. Destabilization of the complex by 1.0 or 1.3 kcal/mol is within the range that would be expected from the elimination of a single nonlinear hydrogen bond. E19A caused a 50-fold loss in binding affinity or a 2.3 kcal/mol destabilization of the complex, consistent with its participation in two hydrogen bonds with RNA (10).

3.2.2 Local cooperative network formed by Asn15, Asn16 and Glu19

To investigate cooperative interactions between Asn15, Asn16, Glu19, we purified a complete set of U1A mutants containing all possible combinations of double and triple Ala substitutions at positions 15, 16 and 19 and measured their affinity for RNA. We applied double mutant cycle analysis (*14*) to assess the coupling energy between these residues (Figure 3-5).

Table 3-1. U1 snRNA hairpin II binding properties of U1A and Ala-substituted U1A mutants

Protein	K_D (M) [a]	ΔG (kcal/mol) [b]	ΔΔG (kcal/mol) [c]	$K_{D\ mut/WT}$ [d]
U1A	$4.9 \pm 0.3\ X10^{-11}$	-14.0 ± 0.04		
N15A	$5.0 \pm 0.8\ X10^{-10}$	-12.7 ± 0.1	1.3	10
N16A	$2.9 \pm 0.4\ X10^{-10}$	-13.0 ± 0.1	1.0	6
E19A	$2.6 \pm 0.2\ X10^{-9}$	-11.7 ± 0.04	2.3	50
E19D	$2.6 \pm 0.1\ X10^{-10}$	-13.2 ± 0.01	0.5	5
N15A/N16A	$2.4 \pm 0.3\ X10^{-9}$	-11.7 ± 0.1	2.3	50
N15A/E19A	$4.3 \pm 0.2\ X10^{-8}$	-10.0 ± 0.2	4.0	880
N16A/E19A	$2.7 \pm 0.5\ X10^{-8}$	-10.2 ± 0.1	3.8	550
N15A/N16A/E19A	$6.0 \pm 0.6\ X10^{-7}$	-8.5 ± 0.1	5.5	12000

[a] K_D is the mean of at least three separate determinations. The standard deviation of these measurements is given. [b] ΔG is the free binding energy of the protein in complex with the U1 snRNA hairpin II calculated by the equation $\Delta G = -RT\ln K_D$, where T=298K. [c] ΔΔG is the change in free binding energy upon the selective Ala substitutions. [d] Fold-reduction of the K_D of the U1A-U1 snRNA hairpin II complex as a result of the indicated mutations.

Ala substitution of Asn15 and Asn16 simultaneously remove two hydrogen bonds from the complex (*10*) and causes a 2.3 kcal/mol loss in the free binding energy. A comparison of this binding energy loss to the sum of the binding energy losses from the two single-substituted N15A and N16A shows that there is no energetic coupling between Asn15 and Asn16 (Table 3-2). This data shows that these two residues interact with RNA independently even though they are close in space in the complex.

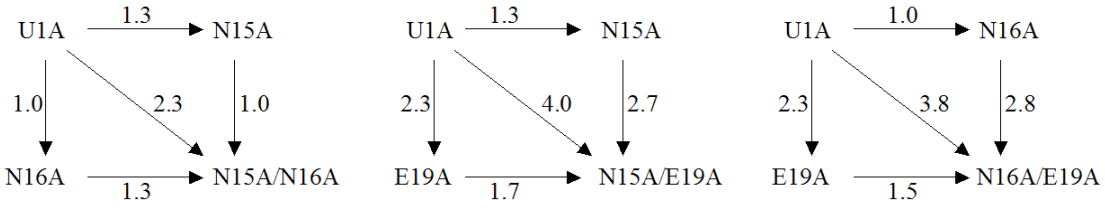


Figure 3-5. Two-dimensional double-mutant cycles of N15A/N16A, N15A/E19A and N16A/E19A. The free energy of binding changes in kcal/mol caused by indicated Ala substitutions are reported in Table 3-1.

Simultaneous Ala substitutions at positions 15 and 19 (N15A/E19A) resulted in a large destabilization of the complex (4.0 kcal/mol), exceeding the sum of the loss of binding energy of the corresponding single mutants (N15A and E19A). This result suggests that Asn15 and Glu19 are energetically coupled upon RNA binding. Double mutant cycle analysis of the residue pair of Asn15 and Glu19 showed a 0.4 kcal/mol energetic coupling between these residues. Substitutions at positions 16 and 19 with Ala (N16A/E19A) destabilized the complex by 3.8 kcal/mol. Using the double mutant cycle analysis, the coupling energy between Asn16 and Glu19 was calculated to be 0.5 kcal/mol (Tabled 3-2).

In order to estimate the energetic coupling of all three residues, we mutated Asn15, Asn16, Glu19 to Ala simultaneously. The complex formed with this triple-substituted U1A mutant was destabilized by 5.5 kcal/mol compared to the wild-type complex (Table 3-1). Asn15, Asn16 and Glu19 individually contributed only limited binding energy to the overall stability of the complex. The sum of these contributions was calculated to be 4.6 kcal/mol, suggesting that the local cooperative network stabilizes the U1A-RNA complex by 0.9 kcal/mol.

Table 3-2. Coupling free energies for N15-N16, N15-E19, N16-E19 and N15-N16-E19

	$\Delta\Delta G_{sum}$ (kcal/mol) [a]	ΔG_{coop} (kcal/mol) [b]
N15-N16	2.3	0
N15-E19	3.6	0.4
N16-E19	3.3	0.5
N15-N16-E19	4.6	0.9

[a] The sum of the changes in free binding energies from the indicated single Ala substitutions. [b] Coupling free binding energies estimated from double-mutant cycles, except for N15-N16-E19, which has been determined by subtraction of the sum of destabilization energies of the three individual Ala substitutions from the overall change upon Ala substitutions of Asn15, Asn16 and Glu19 simultaneously.

While energetic coupling has been often observed for protein-DNA and protein-protein interactions (*27-31*), examples for protein-RNA interactions are rare. The calculated coupling energy from the cooperative network is equal to the sum of individual coupling energies of Asn15-Glu19 (0.4 kcal/mol) and Asn16-Glu19 (0.5 kcal/mol), suggesting that there is no communication between these two couplings (Tabled 3-2). This result is consistent with the previous observation that Asn15 and Asn16 are not coupled with each other, but are coupled with Glu19 in the complex.

Interestingly, we found Asn15, Asn16 and Glu19 together contribute significantly to the overall stability of U1A. This "hot spot" of recognition on the surface of the protein must be built with the help of their interdependent couplings. In comparison with other cooperative networks previously reported for the U1A-RNA complex (*21-23*), the network shown in this study is much more localized, and all of the contacts involved are highly specific. Given the complexity of the binding interface of the U1A-RNA complex, the manner in which individual contacts with RNA and the overall spatial orientations of residues within the hot spot contribute to the complex stability and offers important insight into the energetic origins of complex stability and specificity which will be essential for the design of and inhibition of RNA-RRM complexes.

3.2.3 Phage display of the U1A library

To determine if the identities of Asn15, Asn16, Glu19 in the U1A sequence are required for RNA binding or whether other residues could be functional at these positions, we constructed a combinatorial phage library of the N-terminal RRM of U1A with randomization of positions 15, 16 and 19 and selected for binding to RNA. Phage displayed protein libraries have been widely applied for *in vitro* selections of antibodies, proteins and peptides (*32-34*). The functional N-terminal RRM of U1A has been reported to be successfully displayed on M13 phage with approximately 3% of the expression level (*35-36*). In this study, we chose to use the M13 phagemid vector pCANTAB-5E (see materials and methods) to express the RNA binding RRM of U1A. This construct contains a g3 signal sequence upstream from the U1A insert, which facilitates the expression of the fusion gene, and an E-tag sequence to detect the efficiency of phage displayed libraries. There is an Amber stop codon downstream of the insert, which

facilitates expression of low level fusion proteins inside *E. coli* strain XL1-B (*Sup E*). In order to construct a U1A phage library without any bias towards displaying the wild type U1A protein (*37*), we first constructed an inactive template with stop codon (TAA) insertions at the randomization sites (see materials and methods). The library complexity of genetic codons was simplified by the NN (T/G) scheme. In the native coding sequence of the wild type U1A, Asn is encoded by AAC and Glu is encoded by GAG (*10*). This combination cannot be found in our genetic codon library since, according to our design, Asn is encoded only by an AAT codon. Thus, the AAC codon was used to probe the fidelity of the recovered protein sequences and contamination by the native U1A expressed from the parental phagemid vector. The number of individual clones of the U1A library was determined to be approximately $5x10^7$, which is much larger than the theoretical complexity of this library (20^3). The initial phage stock was denatured and obtained as a single band detected by western blotting (see materials and methods), suggesting the RRMs of U1A variants have been successfully fused to the fd coat protein 3 as designed.

Every member of this protein library may have some affinity for the RNA target in the solution because they probably all maintain the RRM scaffold. In order to select only the tightest binding phage, we applied an RNA concentration of $5x10^{-11}$ M in selection cycles. This concentration is comparable to the K_D value of the wild type complex. It has been shown in previous selection experiments with a smaller U1A library that a low RNA concentration is important for selection of high affinity RNA binders (*36*). In our selections, the RNA target was 5'-end biotinylated and the controls were designed as selections of binding phage from the surface of beads without added RNA in the binding

reactions. The selection conditions were modified so that the recovery rates of phage from the controls were less than 0.01 %. Under our selection conditions, we observed a high enrichment of RNA binding phage from the phage pool after just one round of selection compared to the low recovery rate from the control. Further rounds of selection appeared to yield no improvement in the enrichment of RNA binding phage, suggesting the peak enrichment round was reached (Figure 3-6).

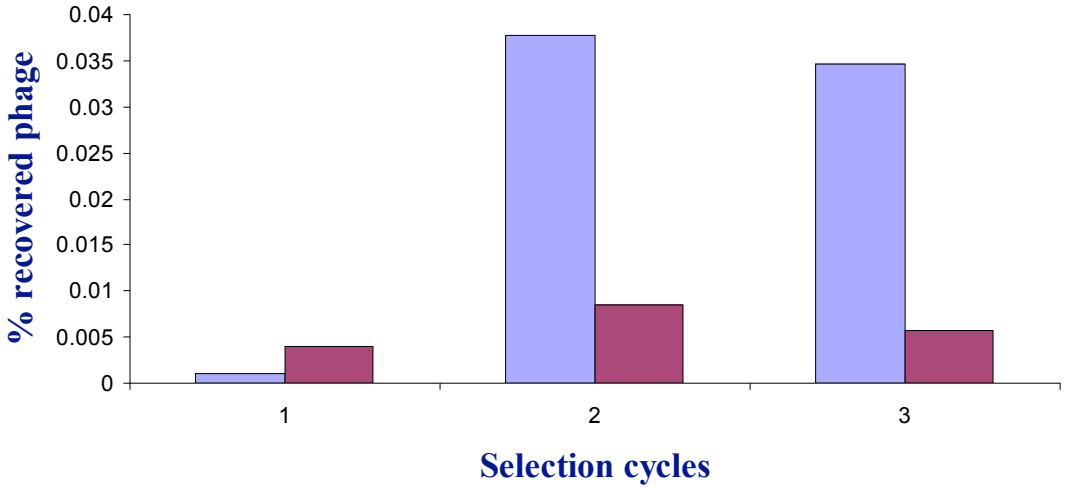


Figure 3-6. Fractions of eluted phage from each selection cycle with U1 snRNA hairpin II as the bait (blue bars) along with eluted phage without the presence of RNA as controls (brown bars). In all cycles of selection, 50 µg of yeast tRNA was applied as the competitor, and RNA binding phage were specifically eluted with 50 µl free RNA for 1hr at room temperature (see materials and methods)

Randomly picked clones from selection cycles 2 and 3 revealed only three selected DNA sequences, possibly due to the small library size and stringent selection conditions we applied. It appeared that no selective bias occurred towards those protein sequences encoded by more than one genetic codon in our library because selected amino acids in all sequenced clones are encoded by only one unique genetic codon. Moreover, we have not found any native U1A coding sequence in those clones (Table 3-3). Taken together,

these results suggest that these sequences were isolated because of their high affinities for RNA under the applied selection conditions instead of the high frequency of occurrence of certain codons or contamination with the U1A wild type sequence.

Table 3-3. Sequences of the selected U1 snRNA hairpin II binding phage.

Selection cycle 2 [a]	Selection cycle 3 [b]
AATAAT---GAG (11) **Asn15Asn16---Glu19**	AATATT---GAG (23) **Asn15Asn16---Glu19**
AATAAT---GAT (2) **Asn15Asn16---Asp19**	CATATT---GAG (1) **His15Asn16---Glu19**

The native sequence of WT U1A (Asn15Asn16---Glu19) is coded for by the DNA sequence AACAAC---GAG in our phagemid vector.
[a] 13 clones randomly picked from plates were sequenced from selection cycle 2. The numbers in parentheses indicate the occurrence of that sequence. The corresponding protein sequences are shown in bold. [b] 24 clones randomly picked from plates were sequenced from selection cycle 3.

The wild type U1A sequence (Asn15Asn16---Glu19) was identified in most of the sequenced clones from selection cycles 2 and 3 (Table 3-3). It was encoded by AATAAT---GAG instead of AACAAC---GAG in the native sequence of U1A. This result suggests these residues are preferred by U1A for RNA binding, and any changes at these positions results in reduced affinity for RNA. Interestingly, there is one U1A mutant E19D that was observed two times among 13 selected clones from selection cycle 2. Asp is found in loop 1 of U2 B" at the same position as Glu19 in the U1A protein (*10*) and may, therefore, contribute to discrimination of the loop IV of U2 RNA from the loop II of U1 RNA. In order to probe the effect of E19D on the stability of the U1A-RNA complex, this mutant was also expressed in *E. coli* and its binding affinity for RNA was

determined (Table 3-1). Not surprisingly, the E19D mutant binds tightly to RNA, with a K_D of 0.26 nM. Although the affinity of E19D for RNA is higher than that of any other U1A mutants tested in this study, it is not as high as that of the wild type U1A protein. In contrast to the effect of E19A, the decrease in the free binding energy upon mutation of Glu19 to Asp was determined to be 0.82 kcal/mol. Asp contains the same carboxyl group on the side chain as Glu, with one fewer methylene group. This result suggests the occupancy of a residue with the carboxyl side chain at position 19 is critical for maintaining the stability of the U1A-RNA complex, and the side chain of Asp is still capable of contacting RNA in a similar geometry as Glu in the wild type U1A. But the distances between hydrogen bonding donors and hydrogen bonding acceptors have not been optimized, possibly due to restrictions caused by the four flanked residues (Figure 3-4). We have observed these residues are helpful for orientating Glu19 outwards for the recognition of G4U RNA (see chapter 4). On the other hand, the shorter side chain of Asp might be beneficial for binding the loop IV of U2 RNA by U2B" (*10*), which has a similar recognition sequence as the loop II of U1 RNA, but has one additional base in the loop region. We propose this spatial arrangement of RNA might prefer a shorter carboxyl group for contacts with U2B" in order to maintain the integrity of the hydrogen bonding interactions. We also found one clone from selection cycle 3 containing His at position 15 as a replacement of Asn in the wild type U1A protein (Table 3-3). This is interesting because His shows a similar arrangement of H-bond donor and acceptor as Asn. The selective advantage of His suggests His may contact RNA with a similar pattern as Asn at position 15 in the U1A protein.

3.3 Conclusions

We have identified a cluster of residues that is comprised of Asn15, Asn16 and Glu19 that cooperatively contributes to the high affinity of the U1A-RNA complex. These positions are occupied by other residues in different members of the RRM family (*38*), suggesting they may be important for the high specificity of the U1A-RNA complex. Selection experiments demonstrate that these amino acids are required for the high affinity of RNA binding. Our analysis also shows that a cluster built by these residues forms a "hot spot" on the surface of U1A. This property could be general for RNA recognition by other RRMs. Considering the similar binding pattern of RNA recognition by RRMs and the widespread of RRMs in a large variety of biological processes, identification of these hot spots formed by energetically coupled residues could be used to rationally design small molecules or peptide mimics as inhibitors for controlling these processes. Our findings complement other investigations of long-range cooperative networks previously reported in the U1A-RNA complex. Taken together, these results define a complex series of cooperative networks involving residues identified by X-ray that make direct contacts with RNA.

3.4 Materials and methods

3.4.1 General procedure for purification of nucleic acids

Oligonucleotides and RNA were purchased from IDT and purified as follows: In a 250 ml clean beaker, the following solutions (National Diagnostics) were mixed: 24 ml of sequaGel concentrate, 43.5 ml sequaGel Dilute and 7.5 ml sequaGel buffer. A 5 ml portion of the gel solution was removed for preparing the plug by addition of 10 µl tetramethylethylenediamine (TEMED) and 20 µl of 10 % ammonium persulfate (APS).

The solution was mixed well and poured quickly into the gap between the glass plates and allowed to polymerize for at least 10 min. The remaining acrylamide solution was used for preparing the resolving gel. Polymerization was initiated by addition of 140 µl of TEMED and 280 µl of 10 % APS. The gel was allowed to polymerize at least one hour before use.

After the polymerization was complete, the comb was removed carefully. The electrophoresis buffer chambers were filled with 1 x TBE buffer and the gel was pre-run for ~1 hr at 50 W. Samples were prepared by mixing equal volumes of 2 x formamide loading buffer (9 ml formamide, 1ml 10 x TBE, 0.0025 g xylene cyanol and 0.0025 g bromophenol blue) and the samples. Just before loading the samples, the wells of the gel were flushed with 1 x TBE to remove urea from the bottom of the wells. Samples were heated for 2 min at 90 ºC and loaded immediately to the wells. The gel was run at 50 W until the bromophenol blue tracking dye reached the bottom quarter of the gel (3-4 hr).

After electrophoresis, the gel was removed from the apparatus, the two plates were removed, and the gel was wrapped with a piece of plastic film. A TLC plate was placed under the gel, and the bands were visualized with a short wave length UV lamp. The bands were cut out with a clean scalpel and placed into autoclaved 2 ml Eppendorf tubes. The gel pieces were crushed to pulp with a 1 ml pipet tip and transferred into a few Eppendorf tubes containing 0.5 ml of 0.3 M sodium acetate (pH 5.2). Tubes were rocked for 10 hr (or over night) at room temperature and spun at full speed in a microcentrifuge for a minute to pellet the acrylamide fragments. 2 volumes of absolute ethanol was added to the supernatant, mixed well and let stand at –20 ºC for at least 2 hr. The solutions were centrifuged at 4 ºC at highest speed for 30 min. The supernatants were decanted and the

pellets were washed twice with 200 µl 70 % ethanol. The pellets were dried by speed vac. Finally, the dried pellets were dissolved in 1 x TE buffer and stored at –20 ºC. The concentrations were measured at 260 nm.

3.4.2 Radiolabeling RNA

In a screw-cap tube, the following solutions were mixed: 5 pmoles of RNA (in <10 µl of solution), 5 µl of [γ-^{32}P] ATP, 2 µl of 10 x kinase buffer, 8 µl of water, and 1 µl of T4 polynucleotide kinase (10,000 units/ml, New England Biolabs, Inc.). The reaction was incubated at 37 ºC for 1.5 hr. After incubation, the reaction was diluted to 50 µl with water and extracted with 50 µl phenol/chloroform/isoamyl alcohol (25:24:1), and 50 µl of chloroform/alcohol (24:1) solution, respectively. 2 µl of 5M NaCl was added to the aqueous RNA solution, followed by 150 µl cold ethanol. The solution was mixed, kept on ice for 15 min and then spun at 4 ºC for 20 min (14k rpm). The supernatant was removed and the pellet was dissolved in 50 µl of water. Ethanol precipitation was repeated. The pellet was dried by removal of water by lyophilization and resuspended in 1 ml of 1 x TE buffer (final concentration: ~5 nM). The radiolabeled RNA was distributed in 10 µl aliquots and stored at –20 ºC.

3.4.3 Isolation of dU-ssDNA template from the phagemid vector pKC

The U1A (2-102) coding sequence was obtained from Nagai (*10*) and inserted into the expression vector pAED 4. This constructed phagemid is called pKC. A His_6 tag was inserted at the N-terminus of the U1A coding sequence in pKC. From a fresh plate (prepared the day before), a single colony of *E. coli* CJ236 (dut-/ung-) harboring the pKC phagemid was picked and put into 1 ml of 2 x YT media supplemented with antibiotics ampicillin (50 µg/ml) and chloramphenicol (5 µg/ml) to maintain the host F'

episome and the phagemid. Pure glucose (20 µl) was added to the culture to a final concentration of 2 % to inhibit transcription of gIII. After shaking at 200 rpm for 7-8 hr at 37 ºC, helper phage M13K07 was added to the culture to a final concentration of 10^{10} phage/ml. Shaking was continued at 200 rpm for an additional 15 min at 37 ºC. The culture was then transferred to 30 ml of 2 x YT/ampicillin/Uridine (0.25 µg/ml) and shaken overnight at 200 rpm at 37 ºC. After centrifugation for 10 min at 15k rpm at 4 ºC, the supernatant was transferred to a new Eppendorf tube containing 6 ml of PEG/NaCl. After incubation at room temperature for 0.5 hr, the phage pellet was collected by centrifugation for 10 min at 10k rpm at 4 ºC.

The phage pellet was resuspended in 0.5 ml of 1 x TBS (50 mM Tris, 150 mM NaCl, pH 7.5) and the solution was spun for 5 min at 15k rpm at 4 ºC to pellet insoluble matter. The supernatant was then transferred to a 1.5 ml tube. 7 µl of buffer MP (Qiagen spin M13 kit) was added and the solution was mixed well. After incubation at room temperature for at least 2 min, the sample was applied to a Qiaprep spin column and centrifuged for 15 s at 8k rpm. The flow-through was discarded and the membrane containing dU-ssDNA was washed and eluted following the recommended protocol by manufacturer. The eluted dU-ssDNA was analyzed by electrophoresing 1 µl on a TBE/agarose gel along with the parental phagemid and a DNA marker (Figure 3-7). The concentration of dU-ssDNA was determined by measuring absorbance at 260 nm (A_{260}=1 for 33 ng/ul of ssDNA).

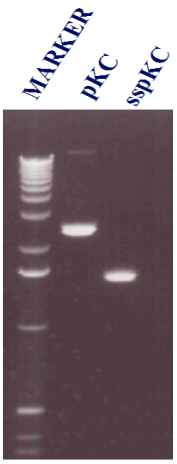


Figure 3-7. Isolation of dU-ssDNA template from M13 phage. The eluted ssDNA were electrophoresed on a 1.0 % TBE/agarose gel containing ethidium bromide for DNA visualization.

3.4.4 Ala mutations of U1A (Kunkel mutagenesis)

Mutations were introduced by standard Kunkel mutagenesis (Figure 3-8) using single oligonucleotides bearing Ala coding sequences (GCC and GCG) in designated positions (*37*). In a 1.5 ml microcentrifuge tube, the following solutions were combined: 30 pmole (~ 0.6 µg) of the mutagenic oligonucleotide, 2.0 µl of 10 x TM buffer (0.5 M Tris, 0.1 M $MgCl_2$, pH 7.5), 4 µl of 10 mM ATP, 1.0 µl 100 mM DTT. Water was added to a total volume of 20 µl. 40 units of T4 polynucleotide kinase were added to the mixture, the solution was mixed well and incubated for 1 hr at 37 ºC. To the 20 µl of phosphorylated reaction mixture, the following solutions were added: 10 pmol (~ 18 µg) of dU-ssDNA template, 25 µl of 10 x TM buffer, and water to a final volume of 250 µl. These conditions provided an oligonucleotide: template molar ratio of 3:1. The mixture was incubated at 90 °C for 2 min, 50-65 °C for 3 min, and room temperature for 5 min. To the annealed oligonucleotide/template mixture, the following solutions were added: 20 µl of 10 mM ATP, 10 µl of 25 mM dNTPs, 15 µl 100 mM DTT, 12 weiss units T4 DNA ligase, and 50 units T7 DNA polymerase. The reaction was incubated at room temperature overnight.

After the overnight incubation, 1.0 ml of buffer QG (Qiagen DNA purification kit) was added to the reaction. The reaction was mixed well and applied to two Qiaspin columns placed in 2 ml microcentrifuge tubes. After centrifugation at 13k rpm for 1 min, 750 µl buffer PE was added to each column. The column was washed one more time and placed into a new 1.5 ml microcentrifuge tube. The heteroduplex DNA was eluted with 35 µl of ultrapure water and analyzed by a TBE/agarose gel (Figure 3-9). 5 µl of the sample was sent out for sequencing at the Core DNA Sequencing Facility at UIUC.

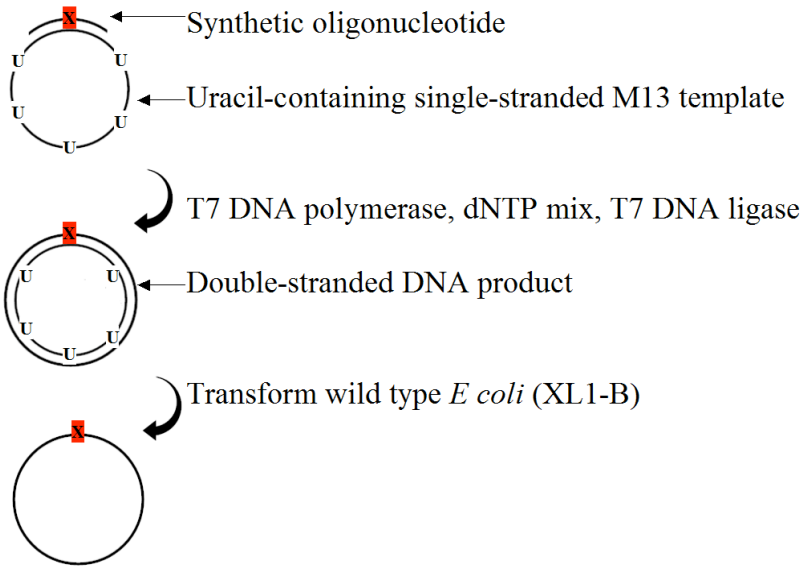


Figure 3-8. Schematic diagram of Kunkel mutagenesis

The following oligonucleotides were used to introduce Ala mutations into the vector:

(N15A) 5'-CGCCCTAACCACACTATTTATATCGCCAACCTCAATGAGAAGATCA AG-3'; (N16A) 5'-CCTAACCACACTATTTATATCAACGCCCTCAATGAGAAGAT CAAGAAGG-3'; (E19A) 5'-TATTTATATCAACAACCTCAATGCGAAGATCAAG AAGGATGAGCTAA-3'; (E19D) 5'-CTATTTATATCAACAACCTCAATGATAGAT

CAAGAAGGATGAGC-3'; (N15A/N16A) 5'-CCACACTATTTATATCGCCGCCCT

CAATGAGAAGATCAAG-3; (N15A/E19A) 5'-CCACACTATTTATATCGCCAACC

TCAATGCGAAGATCAAGAAGGATGAG-3'; (N16A/E19A) 5'-CCACACTATTTAT

ATCAACGCCCTCAATGCGAAGATCAAGAAGGATGAG-3'; (N15A/N16A/E19A)

5'-CCACACTATTTATATCAACGCCCTCAATGCGAAGATCAAGAAGGATGAG-3'

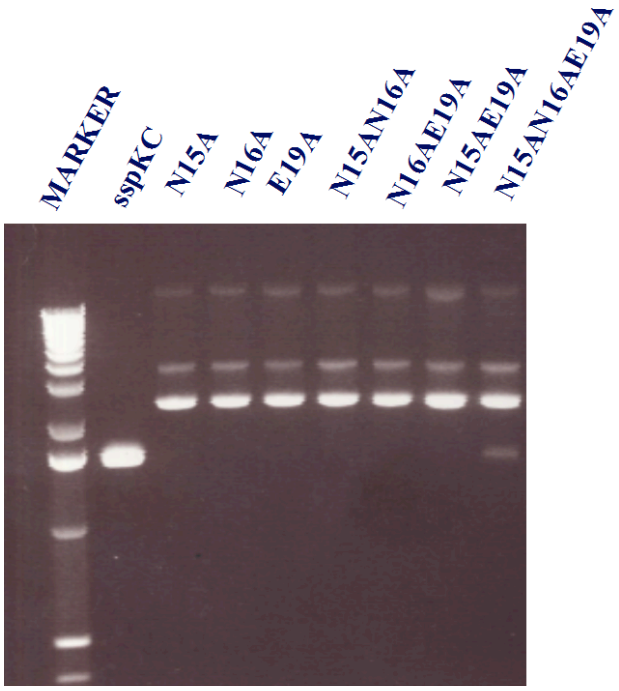


Figure 3-9. Enzymatic synthesis of covalently closed circular, double-stranded DNA (CCC-dsDNA). The eluted heteroduplex DNA was electrophoresed on a 1.0 % TBE/agarose gel containing ethidium bromide for DNA visualization.

3.4.5 Protein expression and purification

Proteins were overexpressed as His_6-tagged fusions in the *E. coli* strain BL21(DE3)pLysS. To a 400 ml culture of log phase *E. coli* cells ($OD_{600} \approx 0.8$) harboring the expression vector was added 1 mM isopropyl 1-thio-P-D-glactopyranoside. The

culture was grown for an additional 3 hr at 37 ºC. The cells were harvested by centrifugation at 4000 g. After freezing overnight at – 80 ºC, the cells were resuspended in 4-5 ml lysis buffer (10 mM imidazole, 500 mM NaCl, 50 mM NaH_2PO_4, 0.5 mM PMSF, pH 7.2) and lysed by ultrasonication with 30 s continuous pulses. The supernatant was isolated by centrifugation at 10000 g for 20 min and was filtered through a 0.45 µM membrane. The filtrate was loaded onto 0.5 ml bed volume of pre-equilibrated Ni^{2+}-agarose beads (Qiagen), the column was washed with gradient buffers seven times (20 mM to 80 mM imidazole, 500 mM NaCl, 50 mM NaH_2PO_4, pH 7.2). The protein was eluted from the column with 1 ml elution buffer (250 mM imidazole, 200 mM NaCl, 50 mM NaH_2PO_4, pH 7.2) and stored at –80 ºC with 20 % glycerol until use. The purity of proteins was estimated by SDS-PAGE (Figure 2-11) to be > 98 % and the identities were confirmed by ESI mass spectrometry (Figure 3-11 to Figure 3-19). All protein concentrations were determined spectrophotometrically with ε_{280} =5190 M^{-1} cm^{-1}.

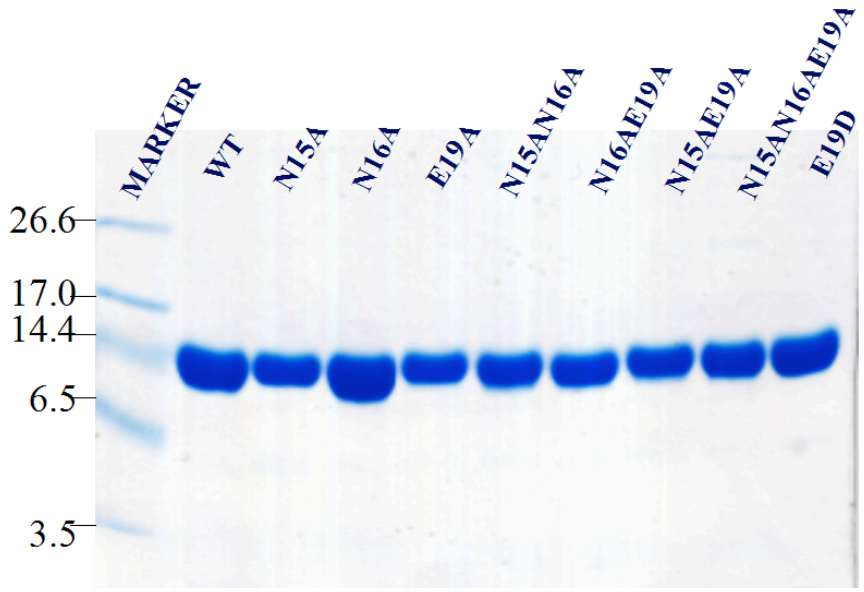


Figure 3-10. Analysis of the purity of the U1A mutants by 12 % SDS-PAGE. Molecular weight markers are in kDa

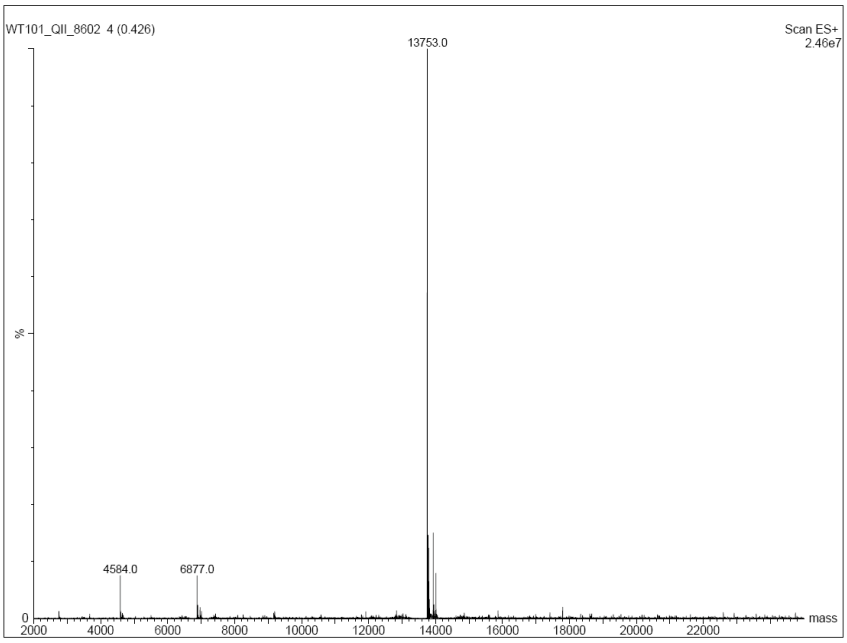


Figure 3-11. Electrospray ionization mass spectrum of the wild-type U1A protein. Calculated MW: 13755 Daltons. Observed MW: 13753 Daltons.

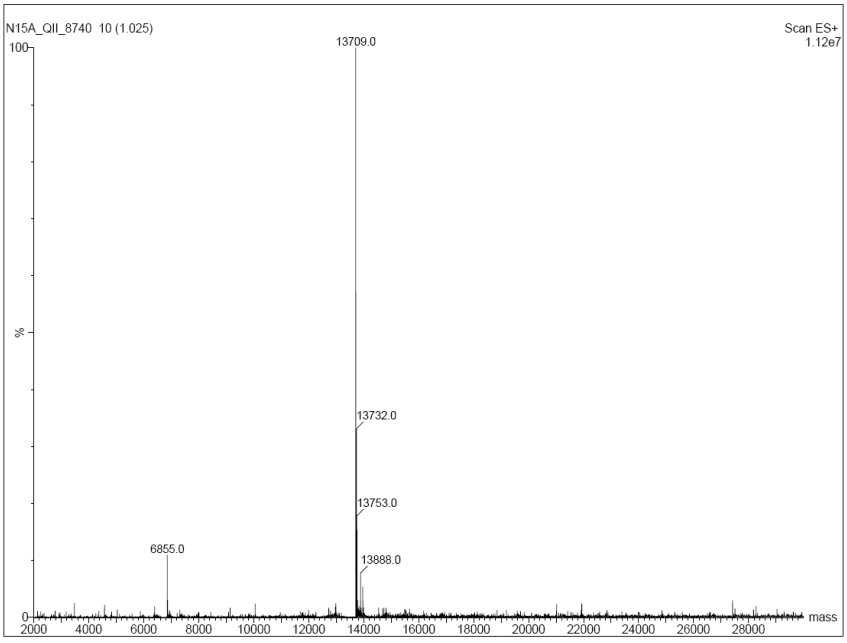


Figure 3-12. Electrospray ionization mass spectrum of the U1A mutant N15A Calculated MW: 13712 Daltons. Observed MW: 13709 Daltons.

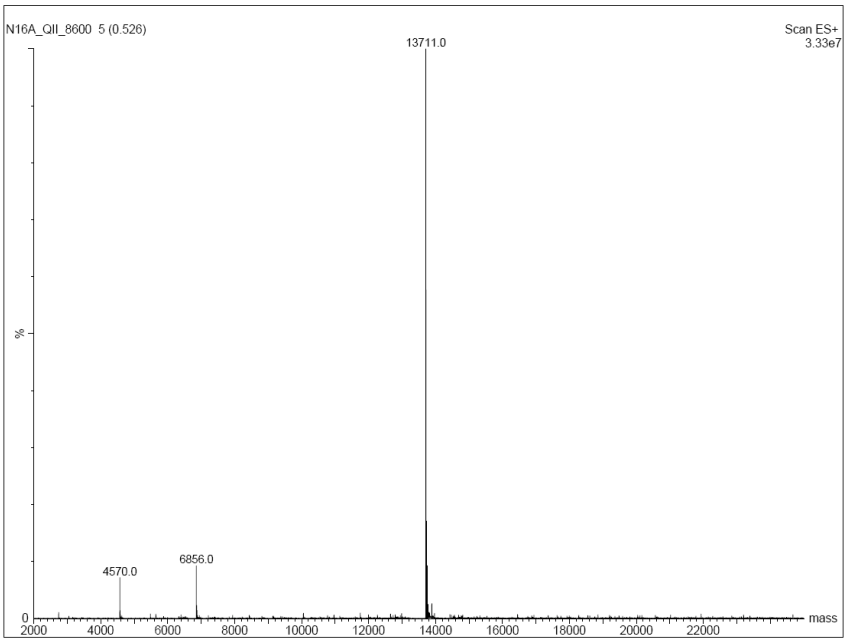


Figure 3-13. Electrospray ionization mass spectrum of the U1A mutant N16A. Calculated MW: 13712 Daltons. Observed MW: 13711 Daltons.

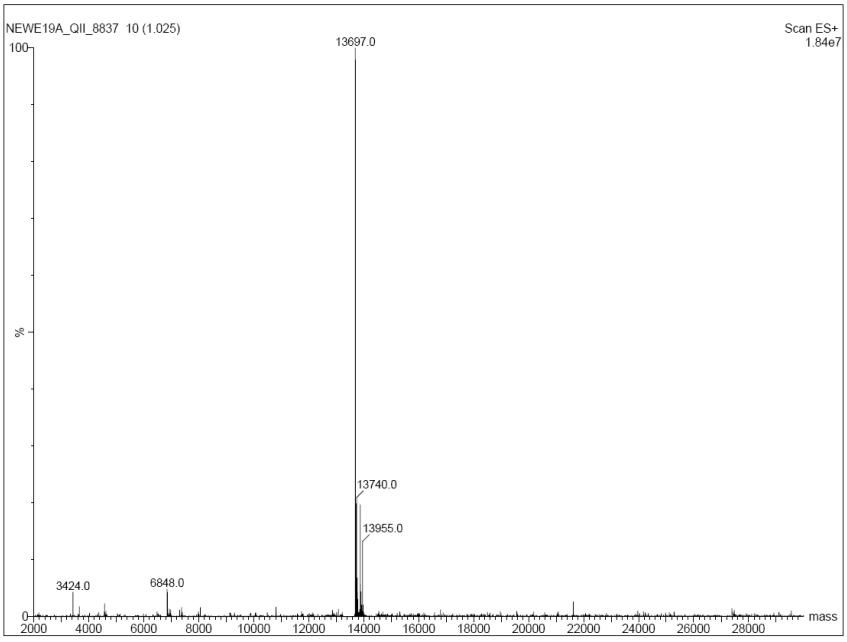


Figure 3-14. Electrospray ionization mass spectrum of the U1A mutant E19A. Calculated MW: 13697 Daltons. Observed MW: 13697 Daltons.

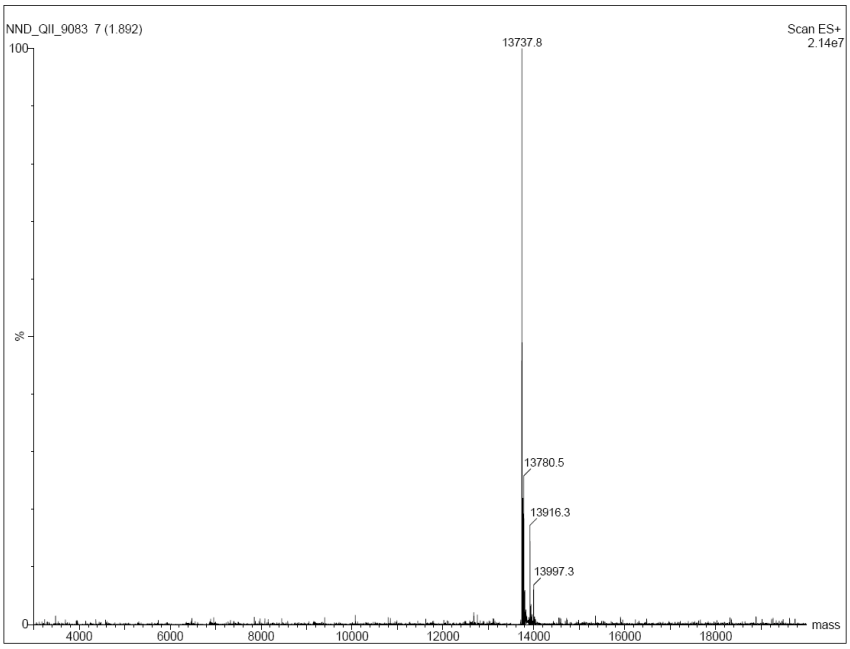


Figure 3-15. Electrospray ionization mass spectrum of the U1A mutant E19D. Calculated MW: 13741 Daltons. Observed MW: 13739 Daltons.

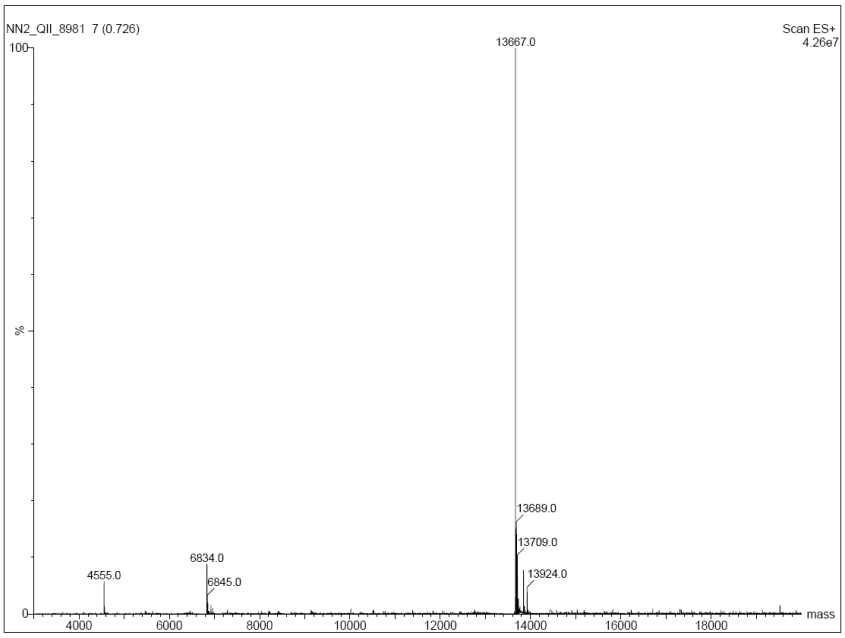


Figure 3-16. Electrospray ionization mass spectrum of the U1A mutant N15A/N16A. Calculated MW: 13669 Daltons. Observed MW: 13667 Daltons.

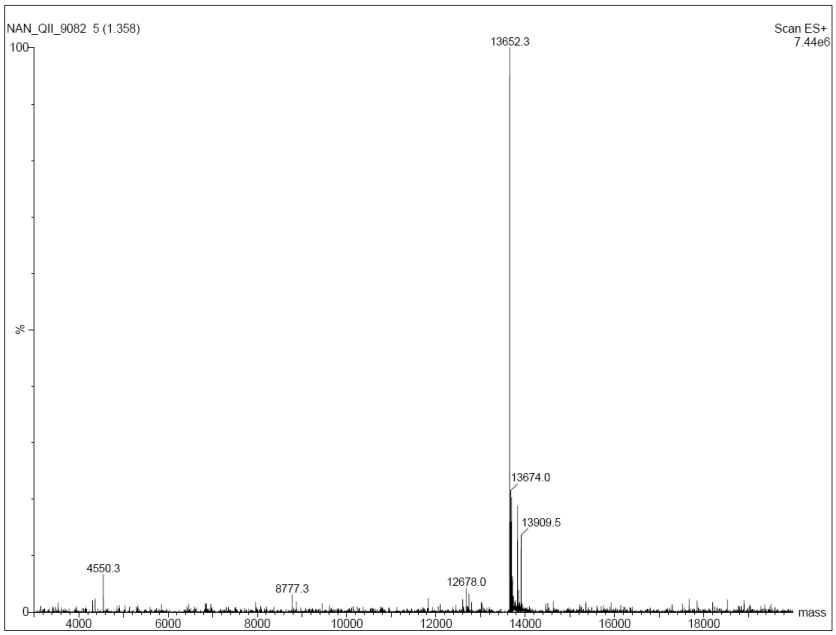


Figure 3-17. Electrospray ionization mass spectrum of the U1A mutant N15A/E19A. Calculated MW: 13654 Daltons. Observed MW: 13652 Daltons.

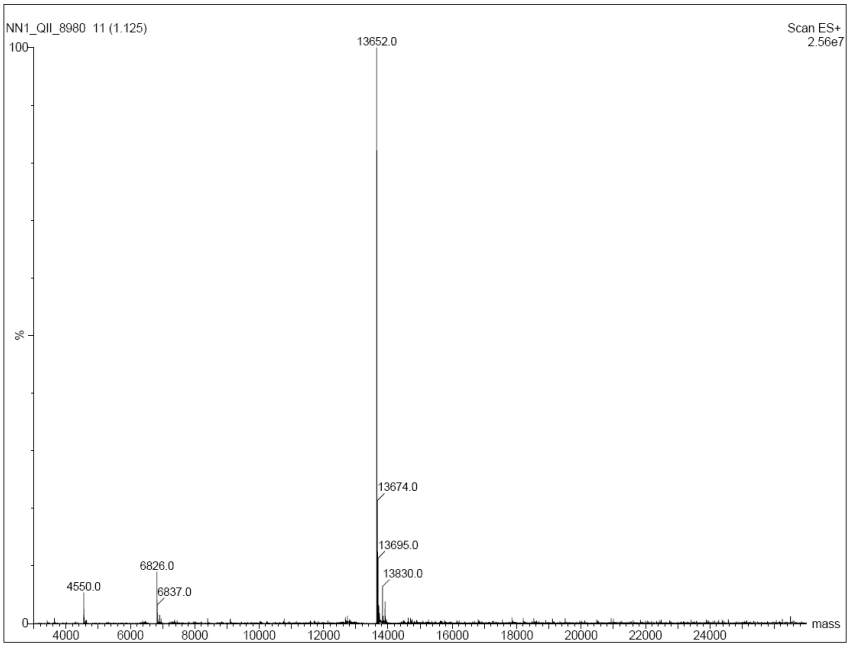


Figure 3-18. Electrospray ionization mass spectrum of the U1A mutant N16A/E19A. Calculated MW: 13654 Daltons. Observed MW: 13652 Daltons.

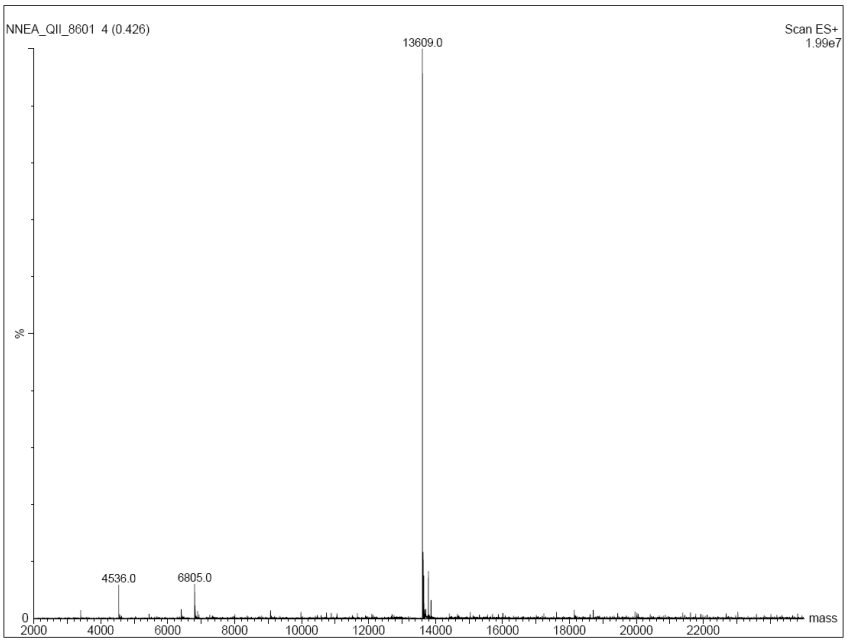


Figure 3-19. Electrospray ionization mass spectrum of the U1A mutant N15A/N16A/E19A. Calculated MW: 13611 Daltons. Observed MW: 13609 Daltons.

3.4.6 Binding affinity assays

Electrophoretic mobility shift assays (EMSA) were used to determine the binding affinity of various U1A proteins for RNA. RNA was purchased from IDT and purified on a 10 % denaturing polyacrylamide gel. The RNA was then 5' end-labeled with [γ-^{32}P] ATP (PerkinElmer). For EMSA, a series of binding reactions comprising a serial dilution of the protein and ^{32}P-labelled RNA (~18 pM) in 10 µl binding buffer (50 mM Tris, 150 mM NaCl, 0.5 mM EDTA, 0.5 % Tween-20, 0.5 mM DTT, 4 % glycerol, 0.5 mg/ml tRNA, pH 7.4) were incubated at room temperature for 30 min. These reactions were then loaded onto an 8 % native polyacrylamide gel and electrophoresed for 30 min at 350 V at 25 ºC. The first lane of the gel was always a control comprising of the labeled RNA alone. Gels were dried and analyzed using a Molecular Dynamics phosphoImager as previously described (*39*). K_Ds were calculated as the concentration of the protein with 50

% of the RNA bound by the software KaleidaGraph (Synergy Software, PA) with the equation: fraction bound $= 1/(1+K_D/[P])$.

3.4.7 Construction of U1A-pIII phagemid library

A monovalent displayed U1A phage library was constructed as a g3p fusion using the Kunkel method (*37*). We applied the genetic code NNM to introduce randomization at target sites. This simplified genetic code reduces amino acid codon numbers and produces much more full-length protein fusion compared to the standard genetic code. Briefly, a synthetic oligonucleotide carrying degenerate codons NNM (N = G, C, A or T, M = G or T) at positions 15, 16 and 19 was annealed to the U-containing silent DNA template (a single-stranded DNA with stop codon TTAs inserted at positions 15, 16, 19). This template was isolated from a pCANTAB 5E phagemid vector (Amersham Pharmacia) containing cDNA of the N terminal RRM of U1A (see materials and methods 3.4.3, Figures 3-20, 3-21).

After extension and ligation reactions, the generated circular form (Figure 2-23) of the new construct was transformed into *E. coli* XL1-Blue cells by electroporation followed by 15 min incubation at 37 ºC. The entire culture was then plated onto LB agar plates supplemented with 100 μg/ml of ampicillin and 2 % glucose. After overnight incubation, the grown clones were pooled together, and a fraction of pooled library containing ~6 x 10^5 clones in 25 μl of LB was amplified once in the presence of M13O7 helper phage at 30 ºC for 18 hr (*36*). The crude phage collected from the supernatant was purified twice by PEG precipitation. The recovered phage pellet was resuspended in 200 μl binding buffer (10 mM Tris, 150 mM NaCl, 0.1 % BSA, 0.5 % Tween-20, pH 7.5) as

the original phage stock. The complexity of the phage library was proven by sequencing 15 individual clones before amplification (sequences are not shown).

Figure 3-20. Construction of the U1A phage library with randomizations at position 15, 16 and 17

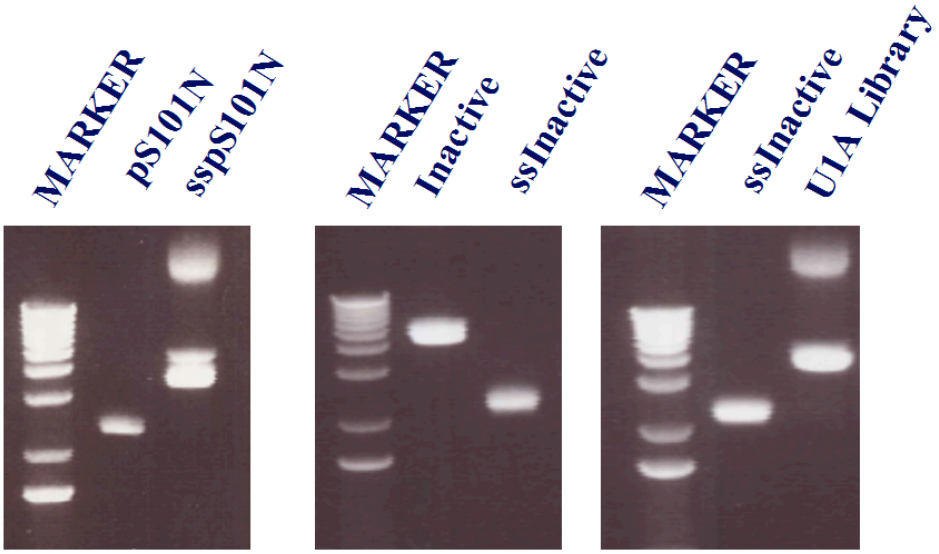


Figure 3-21. *In vitro* synthesis of the U1A phage library. The eluted heteroduplex DNA were electrophoresed on a 1.0 % TBE/agarose gel containing ethidium bromide for DNA visualization. In the inactive phagemid, positions 15, 16 and 19 were replaced by three stop codons in order to terminate the protein expression.

Oligonucleotide sequence for making the inactive template (non-coding strand) (Figure 3-21):

5'ACTTTTTTAGCTCATCCTTCTTGATCTTTTAATTGAGTTATTAGATATAAATA GTGTGGTTAGGGCGGGTCTCGG-3'

Degenerate oligonucleotide sequence for making the U1A phage library (non-coding strand) (Figure 3-21):

5'-ACTTTTTTAGCTCATCCTTCTTGATCTTMNNATTGAGMNNMNNGATATAAA TAGTGTGGTTAGGGCGGGTCTCGG-3'

3.4.8 Detection of the displayed U1A library by western blotting

The identity and expression level of fusion proteins were confirmed by western blotting detecting the E-tag sequence immediately following the coding sequence of U1A in pCANTAB 5E (Figure 3-22). In a 250 ml beaker, the following solutions were mixed: 3.3 ml of 22.2 % acrylamide/0.6 % bisacrylamide, 3.8 ml of 1 M Tris-HCl (pH 8.8), 2.5 ml of water, 100 µl of 10 % SDS, 50 µl of 10 % APS, and 5 µl of TEMED. The mixture was immediately poured into the gel apparatus, overlaid with about 200 µl of water and allowed to polymerize for about 30 min. Meanwhile, the stacking gel mix was prepared: 1 ml of 22.2 % acrylamide/0.6 % bisacrylamide, 0.6 ml of 1 M Tris-HCl (pH 6.8), 3.3 ml of water, 50 µl of 10 % SDS, 25 µl of 10 % APS, and 5 µl of TEMED. The water was poured off of the gel before the stacking solution was poured onto the top of the gel, and the comb was inserted. The gel was allowed to polymerize for 30 min. 10 µl of 2 x SDS loading buffer was added to 10 µl of the phage stock, the mixture was heated at 95 ºC for 5 min, and loaded onto the gel immediately. A pre-stained molecular weight marker was loaded into a neighboring well. The gel was run at 120 V until the blue dye in the loading

buffer reached the bottom of the gel. The proteins were then transferred to a pre-cut nitrocellulose membrane (0.45 µm, Schleicher & Schuell) in an electroblotting device (Trans-blot SD cell, Bio-rad) for 1 hr at 20 V as recommended by the manufacturer. After the transfer was complete, the membrane was blocked at least 1 hr with TBST buffer (50 mM Tris, 150 mM NaCl, 0.1 % TW-20, 5 % milk, pH 7.5). After three washes, the HRP conjugated E-tag antibody (1: 5000 in TBST) was added to the membrane and incubated for 1 hr at room temperature. After three washes, the membrane was developed with 1-Step Ultra TMB-ELISA (Pierce) for 15 min at room temperature.

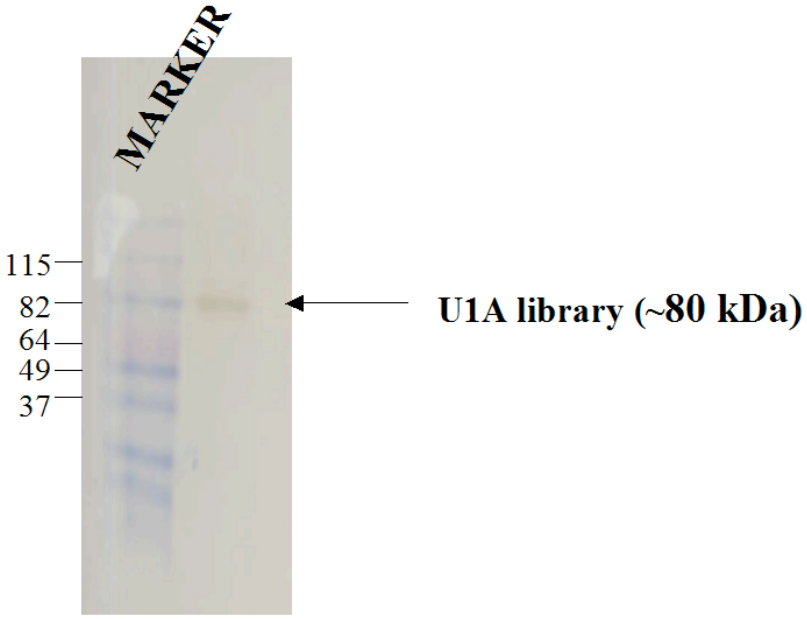


Figure 3-22. Western blotting analysis of the displayed U1A phage library. 20 µl of phage solution was loaded on a SDS-PAGE (8%). The blot was probed with an HRP-conjugated antibody against the E-tag. The band marked with the arrow represents the full-length fusion protein containing the N-terminal RRM of U1A. Molecular weight makers are in kDa.

3.4.9 Phage display selection *in vitro*

Panning selections were performed on streptavidin-coated beads (Dynabeads M-280, Invitrogen) in an indirect capture manner. Briefly, 5'-Biotinylated 25-mer synthetic RNA

was purchased from IDT and purified by electrophoresis on a 10 % denaturing polyacrylamide gel. For every round of selection, 25 µl binding solution containing approximately 5.0 x 10^{11} phage, 2 fmole 5'- biotinylated RNA, 50 µg of yeast tRNA and 20 units of Rnasin in binding buffer (10 mM Tris, 150 mM NaCl, 0.1 % BSA, 0.5 % Tween-20, pH 7.5) were incubated for 30 min at room temperature. The binding reaction was then transferred to 5 µl pre-washed streptavidin-coated beads and incubated for an additional 30 min with gentle shaking. The beads were washed 2-5 times with 1 ml cold binding buffer. The bound phage was eluted by incubation with 50 µl unlabeled RNA for 1 hr at room temperature. The eluate was used to infect the log phase XL1-Blue cells by adding to 1 ml of the cells immediately for 20 min at 37 ºC. The phage were amplified as before and used immediately for the next round of selection. After 3 rounds of selection, individual clones randomly picked from plates were sent for sequencing at the Core DNA Sequencing Facility at UIUC.

3.5 References

1. Perron, M. P., and Provost, P. (2009) Protein components of the microRNA pathway and human diseases, *Methods Mol Biol 487*, 369-385.

2. Repoila, F., and Darfeuille, F. (2009) Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects, *Biol Cell 101*, 117-131.

3. Waters, L. S., and Storz, G. (2009) Regulatory RNAs in bacteria, *Cell 136*, 615-628.

4. Brown, R. S. (2005) Zinc finger proteins: getting a grip on RNA, *Curr Opin Struct Biol 15*, 94-98.

5. Chang, K. Y., and Ramos, A. (2005) The double-stranded RNA-binding motif, a versatile macromolecular docking platform, *FEBS J 272*, 2109-2117.

6. Clery, A., Blatter, M., and Allain, F. H. (2008) RNA recognition motifs: boring? Not quite, *Curr Opin Struct Biol 18*, 290-298.

7. Lin, S., and Fu, X. D. (2007) SR proteins and related factors in alternative splicing, *Adv Exp Med Biol 623*, 107-122.

8. Perez-Arellano, I., Gallego, J., and Cervera, J. (2007) The PUA domain - a structural and functional overview, *FEBS J 274*, 4972-4984.

9. Chen, Y., and Varani, G. (2005) Protein families and RNA recognition, *FEBS J 272*, 2088-2097.

10. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994) Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin, *Nature 372*, 432-438.

11. Varani, G., and Nagai, K. (1998) RNA recognition by RNP proteins during RNA processing, *Annu Rev Biophys Biomol Struct 27*, 407-445.

12. Burd, C. G., and Dreyfuss, G. (1994) Conserved structures and diversity of functions of RNA-binding proteins, *Science 265*, 615-621.

13. Di Cera, E. (1990) Thermodynamics of local linkage effects. Contracted partition functions and the analysis of site-specific energetics, *Biophys Chem 37*, 147-164.

14. Di Cera, E. (1998) Site-Specific Thermodynamics: Understanding Cooperativity in Molecular Recognition, *Chem Rev 98*, 1563-1592.

15. Pineda, A. O., Cantwell, A. M., Bush, L. A., Rose, T., and Di Cera, E. (2002) The thrombin epitope recognizing thrombomodulin is a highly cooperative hot spot in exosite I, *J Biol Chem 277*, 32015-32019.

16. Yang, J., Swaminathan, C. P., Huang, Y., Guan, R., Cho, S., Kieke, M. C., Kranz, D. M., Mariuzza, R. A., and Sundberg, E. J. (2003) Dissecting cooperative and additive binding energetics in the affinity maturation pathway of a protein-protein interface, *J Biol Chem 278*, 50412-50421.

17. Shapiro, R., Ruiz-Gutierrez, M., and Chen, C. Z. (2000) Analysis of the interactions of human ribonuclease inhibitor with angiogenin and ribonuclease A by mutagenesis: importance of inhibitor residues inside versus outside the C-terminal "hot spot", *J Mol Biol 302*, 497-519.

18. Stark, H., Dube, P., Luhrmann, R., and Kastner, B. (2001) Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle, *Nature 409*, 539-542.

19. Law, M. J., Chambers, E. J., Katsamba, P. S., Haworth, I. S., and Laird-Offringa, I. A. (2005) Kinetic analysis of the role of the tyrosine 13, phenylalanine 56 and glutamine 54 network in the U1A/U1 hairpin II interaction, *Nucleic Acids Res 33*, 2917-2928.

20. Nagai, K., Oubridge, C., Ito, N., Jessen, T. H., Avis, J., and Evans, P. (1995) Crystal structure of the U1A spliceosomal protein complexed with its cognate RNA hairpin, *Nucleic Acids Symp Ser*, 1-2.

21. Kranz, J. K., and Hall, K. B. (1998) RNA binding mediates the local cooperativity between the beta-sheet and the C-terminal tail of the human U1A RBD1 protein, *J Mol Biol 275*, 465-481.

22. Kranz, J. K., and Hall, K. B. (1999) RNA recognition by the human U1A protein is mediated by a network of local cooperative interactions that create the optimal binding surface, *J Mol Biol 285*, 215-231.

23. Showalter, S. A., and Hall, K. B. (2002) A functional role for correlated motion in the N-terminal RNA-binding domain of human U1A protein, *J Mol Biol 322*, 533-542.

24. Kormos, B. L., Baranger, A. M., and Beveridge, D. L. (2007) A study of collective atomic fluctuations and cooperativity in the U1A-RNA complex based on molecular dynamics simulations, *J Struct Biol 157*, 500-513.

25. Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C., and Nagai, K. (1991) Identification of molecular contacts between the U1 A small nuclear ribonucleoprotein and U1 RNA, *EMBO J 10*, 3447-3456.

26. Smith, G. P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface, *Science 228*, 1315-1317.

27. Bingle, L. E., Rajasekar, K. V., Muntaha, S., Nadella, V., Hyde, E. I., and Thomas, C. M. (2008) A single aromatic residue in transcriptional repressor protein KorA is critical for cooperativity with its co-regulator KorB, *Mol Microbiol 70*, 1502-1514.

28. GuhaThakurta, D., and Stormo, G. D. (2001) Identifying target sites for cooperatively binding factors, *Bioinformatics 17*, 608-621.

29. Heneghan, A. F., Connaghan-Jones, K. D., Miura, M. T., and Bain, D. L. (2006) Cooperative DNA binding by the B-isoform of human progesterone receptor: thermodynamic analysis reveals strongly favorable and unfavorable contributions to assembly, *Biochemistry 45*, 3285-3296.

30. Moza, B., Buonpane, R. A., Zhu, P., Herfst, C. A., Rahman, A. K., McCormick, J. K., Kranz, D. M., and Sundberg, E. J. (2006) Long-range cooperative binding effects in a T cell receptor variable domain, *Proc Natl Acad Sci U S A 103*, 9867-9872.

31. Senear, D. F., Ross, J. B., and Laue, T. M. (1998) Analysis of protein and DNA-mediated contributions to cooperative assembly of protein-DNA complexes, *Methods 16*, 3-20.

32. Nieri, P., Donadio, E., Rossi, S., Adinolfi, B., and Podesta, A. (2009) Antibodies for therapeutic uses and the evolution of biotechniques, *Curr Med Chem 16*, 753-779.

33. Funke, S. A., and Willbold, D. (2009) Mirror image phage display--a method to generate D-peptide ligands for use in diagnostic or therapeutical applications, *Mol Biosyst 5*, 783-786.

34. Majidi, J., Barar, J., Baradaran, B., Abdolalizadeh, J., and Omidi, Y. (2009) Target therapy of cancer: Implementation of monoclonal antibodies and nanobodies, *Hum Antibodies 18*, 81-100.

35. Laird-Offringa, I. A., and Belasco, J. G. (1995) Analysis of RNA-binding proteins by in vitro genetic selection: identification of an amino acid residue important for locking U1A onto its RNA target, *Proc Natl Acad Sci U S A 92*, 11859-11863.

36. Laird-Offringa, I. A., and Belasco, J. G. (1996) In vitro genetic analysis of RNA-binding proteins using phage display libraries, *Methods Enzymol 267*, 149-168.

37. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection, *Methods Enzymol 154*, 367-382.

38. Birney, E., Kumar, S., and Krainer, A. R. (1993) Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors, *Nucleic Acids Res 21*, 5803-5816.

39. Zhao, Y., and Baranger, A. M. (2003) Design of an adenosine analogue that selectively improves the affinity of a mutant U1A protein for RNA, *J Am Chem Soc 125*, 2480-2488.

CHAPTER 4

SINGLE AMINO ACID SUBSTITUTIONS ALTER THE BINDING PROPERTIES OF THE HUMAN U1A PROTEIN

4.1 Introduction

RNA binding proteins have developed a large diversity of functions through millions of years of evolution (*1-6*). However, most of them are constructed from only a few modular structures defined by different motifs (see Chapter 1). Single or multiple copies of these motifs create versatile binding surfaces through unique structural arrangements that allow specific recognition of different RNA structures and sequences (*7*). The RNA recognition motif (RRM) or ribonucleoprotein (RNP) is one of the largest and best characterized RNA binding protein families. More than 6000 RRMs have been identified in proteins involved in the regulation of diverse post-transcriptional activities, such as pre-mRNA splicing, editing, RNA stability and transport (*4, 8*). The RRM is usually 80-90 residues long and consists of a conserved α/β fold featuring a four-stranded anti-parallel β-sheet and two α-helices. The solvent-exposed β-sheet serves as the major binding platform for RNA recognition. Several highly conserved residues that interact with RNA have been identified in the β-sheet, giving the basal affinity of RNA binding (*9-10*). For example, Arg and Lys residues are usually found in the two conserved segments of the two central β strands in the β-sheet. These residues are capable of participating in electrostatic interactions with the phosphodiester backbone of RNA; Tyr and Phe are also conserved in these two β strands, and both make stacking interactions with RNA bases. All of these interactions are thought to be important to maintain the

stability of the RRM-RNA complex and help position the RNA in optimal orientations for interactions with other residues in the protein.

Although proteins containing RRMs use a conserved recognition surface to bind RNA, they have evolved to bind RNA with different specificities to achieve their designated functions (*7*). We reasoned that the different specificities could be partially due to residues that make base-specific contacts, but are not well conserved in the RRM family. We have chosen the N-terminal RRM of the U1A protein as our model system to explore the possibility of building RNA binding proteins with altered RNA specificity because U1A is well characterized and the interactions of U1A in complex with RNA are known in detail (*11*). The N-terminal RRM of the human U1A protein binds to a single-stranded RNA element AUUGCAC in the U1 snRNA hairpin II loop with high affinity and specificity. A similar pattern has also been found for U1A binding to the 3'-untranslated region (3'UTR) of its own pre-mRNA (*12-13*). The crystal structure of the complex with the U1 snRNA hairpin II loop has revealed a flat and complementary binding interface between U1A and RNA. In the complex, every base in the recognition element of RNA splays out and contacts its distinct recognition region of the protein, mainly via arrangements of hydrogen bonds and conserved stacking interactions (*11*).

Surprisingly, only a small set of interactions was identified on the binding interface to involve amino acid side chains of the U1A protein. The recognition of G4 in the AUUGCAC sequence of RNA appears crucial because it makes the most extensive hydrogen bonding interactions with the protein. The substitutions of G4 with other natural bases dramatically decrease the affinity of the complex by at least 10,000-fold (*14-16*). We have shown previously that the local cooperativity of Asn15, Asn16 and

Glu19 regulates the high affinity of U1A-RNA (see Chapter 3). All of these residues are involved in the recognition of G4 in the U1A-RNA complex (Figure 4-1). For example, the amide nitrogen of the side chain of Asn15 forms a specific hydrogen bond with N7 of G4. One oxygen of the carboxylate group of Glu19 forms a specific hydrogen bond with the 2-NH_2 of G4, while the other oxygen contacts the 3-NH of U3. The side chain of Asn16 has no contact with G4, it forms a specific hydrogen bond with U3 of the RNA. But its presence at position 16 appears important for the rigidity of the backbone of the protein in this region, because there is no energetic coupling between Asn15 and Asn16 in the wild type U1A protein (see Chapter 3). This could be crucial for the formation of two hydrogen bonds between functional groups in the main chain of the U1A protein and O6 of G4 (Figure 4-1).

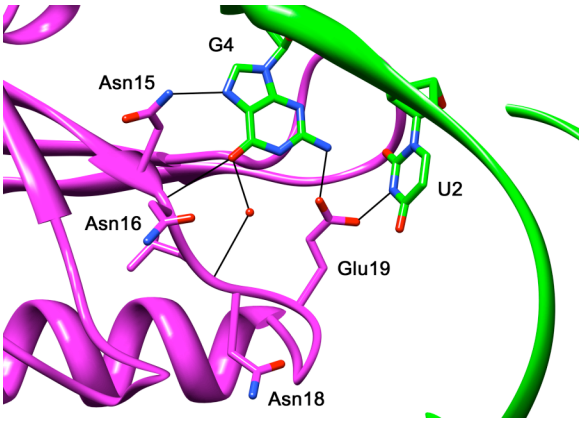


Figure 4-1. Structural representation of hydrogen bonding interactions between Asn 15, Asn 16 and Glu 19 of U1A and G4 of U1 snRNA hairpin II. The hydrogen bonding interactions are shown by solid lines.

Given the specific contacts and spatial arrangements of residues involved in the recognition of G4, we wondered if the disruption of the binding interface due to mutations at G4 of the cognate RNA could be restored by selective mutations of the U1A protein. We explored this question by performing a phage display selection of a U1A

library with randomization at positions 15, 16 and 19, which are the residues (Asn15, Asn16, Glu19) involved in the recognition of G4 in the wild type U1A-RNA complex. The RNA targets for phage selection were G4A RNA, G4U RNA and G4C RNA.

4.2 Results and discussion

The use of the modular structure of a highly conserved protein family to design proteins with novel binding properties attracts great interests in the field of biomedical research (*17-21*). The N-terminal RRM of U1A represents one of the most abundant RNA binding motifs, and it binds to its cognate RNA specifically using the conserved binding surface. On the flat and complementary binding interface of the U1A-RNA complex, there are numerous interactions that are involved in the recognition of the cognate RNA (*11*). We are interested in G4 recognition because it involves two base contacts with side chains of amino acids in the U1A protein. Selective mutations of two base recognition residues have been shown to efficiently alter the specificity of Pumilio repeats (*22*). To isolate U1A mutants with desired affinities for RNA containing G4A, G4C or G4U mutations, we applied the U1A phage library constructed previously (see Chapter 2) for selecting binding phage for these RNA mutants. In this library, Asn15 and Glu19 have been randomized. These residues make specific hydrogen bonding contacts with G4 in the wild type complex. Position 16 was also randomized which allows subtle structural recognition of the backbone of the protein in this region. This could be beneficial for recognition of other bases at the G4 position since it permits optimal orientation of selected residue at position 15 for easy access. In the wild type U1A-RNA complex, O6 in the carbonyl group of G4 forms two hydrogen bonding interactions (direct and water-mediated) with the backbone of Asn16 and Leu17, which further

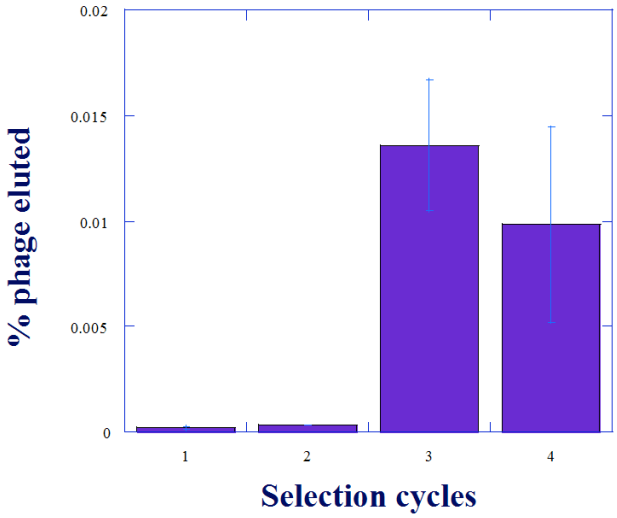
stabilize the U1A-RNA complex (*11*). Structural disturbance of the backbone in this region would also allow possible non-specific contacts with the new base at G4 position.

4.2.1 Phage display selections and ELISA

Phage display selections were performed by an indirect capture method in order to precisely control the desired affinity with RNA sequences bearing G4U, G4A or G4U substitutions (*23*). The eluted U1A phage from selection cycles were monitored by comparing the recovery percentage from each selection. A control experiment without the added RNA target in the binding reaction was not performed as previously because the selection conditions have been optimized so that the U1A phage library has a low retention rate to the surface of the beads (see Chapter 2). Bound phage were eluted with unlabelled RNA to minimize the recovery of non-specifically bound phage. This is crucial for the initial selection, because even a small portion of these background binding phage will be co-amplified along with "authentic" binders. These "false positives" will increase the complexity of the amplified phage library and interfere with the following rounds of selections. We failed to recover the wild type U1A phage using glycine elution buffer (0.2 M glycine-HCl, pH 2.2, 1 mg/ml BSA) from the immobilized wild type RNA, even though the wild type U1A phage showed to be dominant in the library (data not shown), suggesting a specific elution buffer is required for efficient selections. Using ≈ 2 fmol of biotinalyted RNA as the bait for the initial selection, we only observed enrichment of G4U binding phage in the first three rounds of selection (Figure 4-2). We were not able to enrich any binding proteins from selections using G4A or G4C RNA as targets. The failure of recovery of G4A or G4C binding proteins may result from the poor flexibility of the randomized position and the "hard to access" hydrogen bond donors and

acceptors due to the spatial orientations of A or C bases in the mutated RNA, implying a limited adaptability of U1A at these positions.

A



B

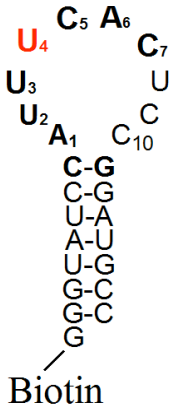


Figure 4-2. (A). Fractions of eluted phage from each selection cycle with the G4U RNA mutant as the bait (blue bars). Selections were repeated twice. In all cycles of selection, 50 µg of yeast tRNA was applied as the competitor, and RNA binding phage were specifically eluted with 50 µl free G4U RNA for 1h at room temperature (see materials and methods). (B). Secondary structure of U1 snRNA hairpin II mutant (G4U) applied in the phage selections. G4 was substituted by U (highlighted in red) in the RNA sequence and the mutated RNA was biotinylated at the 5'-end. The recognition sequence is shown in boldface.

We sequenced nine clones from the enriched pool from the third panning for G4U as the target (Table 4-1). All selected sequences have Asn at position 16, consistent with its role forming a specific base contact with U3, which is intact in the G4U RNA. Asn is preferred at position 15 for G4U binding since it was found seven times in nine sequenced clones. No specific residue preferences were observed for position 19. In order to isolate phage with higher affinity for the G4U RNA target, we applied a 10-fold lower amount of G4U RNA (≈ 0.2 fmol) in the fourth round panning. As a result of this

increased selection pressure, only two U1A mutants with a single-point mutation at position 19, either E19Q or E19H were recovered. Asn residues were retained at both positions 15 and 16 in the sequenced clones from the fourth round of panning.

Table 4-1. Sequences of G4U RNA binding phage.

Selection cycle 3 [a]	Selection cycle 4 [b]
Asn15Asn16---His19 (3)	**Asn15Asn16---Gln19 (13)**
Asn15Asn16---Gln19 (1)	
Asn15Asn16---Lys19 (1)	
Tyr15Asn16---Gln19 (1)	**Asn15Asn16---His19 (3)**
Lys15Asn16---Asn19 (1)	
Asn15Asn16---Thr19 (1)	
Asn15Asn16---Pro19 (1)	

The native sequence of the wild-type U1A is Asn15Asn16---Glu19
[a] 9 clones randomly picked from plates were sequenced from selection cycle 3. The numbers in parentheses indicate the occurrence of that indicated sequence. The corresponding protein sequences are shown in bold. [b] 16 clones randomly picked from plates were sequenced from selection cycle 4.

These two selected U1A mutants, E19Q and E19H were subjected to phage-based ELISA with a microtier plate (streptavidin-coated) pre-coated with the biotinylated G4U RNA. The wild-type U1A phage showed essentially no binding to G4U RNA under applied conditions. E19Q and E19H phage exhibited dramatically greater affinity to G4U RNA compared to the wild type U1A phage, as shown by the intense absorbance at 450 nm (Figure 4-3). These observations qualitatively validate the results from the panning experiments.

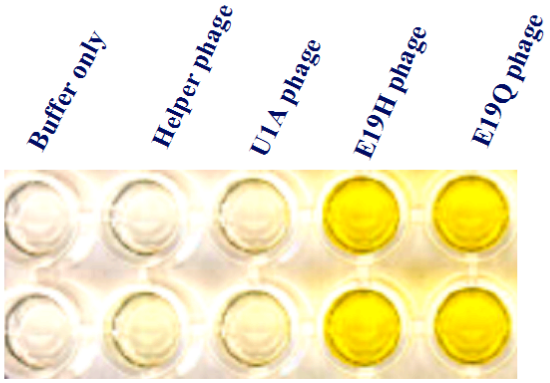


Figure 4-3. Phage-ELISA assays of the isolated G4U binding phage. Assays were performed in duplicate. All wells were coated with the biotinylated G4U RNA (5 μg/ml, 50 μl/well), 100 μl anti-M13 HRP-conjugated antibody (1:5000 in wash buffer) was applied to each well. Assays were developed with 1-Step Ultra TMB-ELISA (Pierce) (see materials and methods).

4.2.2 Binding properties of selected U1A mutants

To evaluate whether the phage display experiments selected U1A mutants with higher affinity for G4U RNA than the wild type U1A protein, we expressed E19Q and E19H U1A mutants (Figure 4-4 and Table 4-2). SDS-PAGE analysis confirmed the high purity of the protein preparations, and the identities of proteins were confirmed by ESI mass spectrometry (see materials and methods).

The wild-type U1A protein binds weakly affinity to G4U RNA (K_D > 2.1 μM). Compared to the wild type RNA, the G4U mutation destabilizes the complex by at least 6.4 kcal/mol in binding energy. E19Q and E19H U1A mutants enriched from the rounds of phage selection binds to G4U 50-fold and 30-fold more tightly than does the wild-type U1A, as shown by 2.3 kcal/mol and 2 kcal/mol increases in binding energy, respectively. Compared to Glu, the side chain of His is more constrained, and this could be responsible for the observed higher affinity of E19Q compared to E19H to G4U RNA.

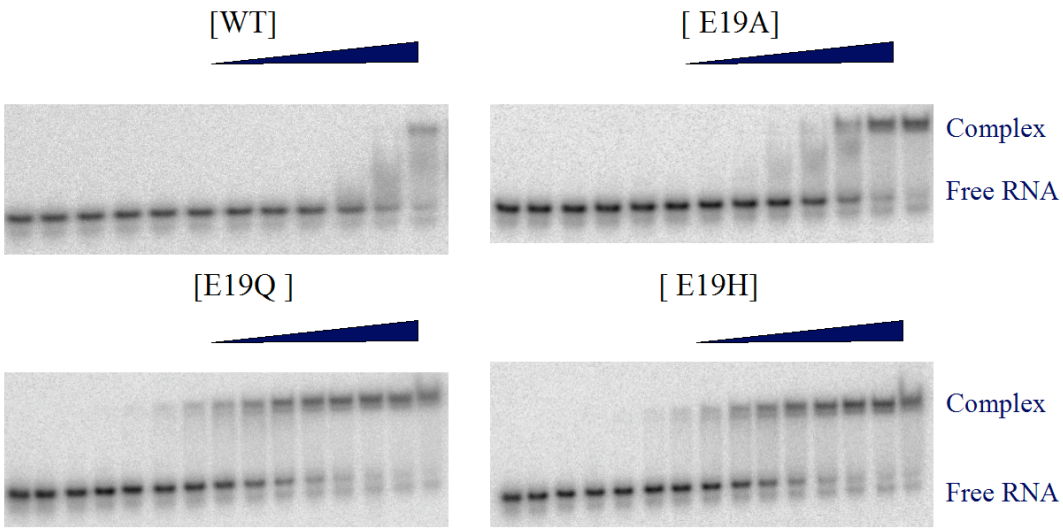


Figure 4-4. Representative electrophoretic mobility shift assays of U1A and its mutants in complex with G4U RNA. Samples from a 4-fold serial dilution of U1A were incubated with ^{32}P-labeled G4U RNA at room temperature for 30 min. The shifted bands were resolved by electrophoresis on an 8 % native polyacrylamide gel. Control reactions without added protein were loaded in the first lane on the left side of each gel.

Asn15 and Asn16 are located at the tip of the β1 strand of the U1A protein (*11*). Selections of both residues suggest that together they can tolerate U at the G4 position better than any other combinations of residues at positions 15 and 16. They may even make new contacts with the mutated G4U RNA. In the U1A-RNA complex, besides the contact between the protein backbone and G4 of the RNA, the side chain amino group of Asn16 also forms a hydrogen bond with N3 of U3 of the wild type RNA, which was kept intact in the mutated G4U RNA. This original contact should be maintained by the selection of Asn at position 16. Among all 20 natural amino acids, Gln contains the same functional side chain as Asn, only with one additional methylene group longer. It was not selected at position 16, suggesting that the integrity of the spatial orientation of the mutated RNA loop was maintained in the mutant complex. The hydrogen bond between Asn16 and N3 of RNA U3 contributes no more than 1 kcal/mol of binding energy in the

wild type U1A-RNA complex (see Chapter 3), its selection at position 16 from phage pannings suggests our customized selection protocol is sensitive to small binding energy changes, and the randomization at position 16 may serve as an internal control of other U1A libraries for selection of mutated RNA binding proteins, providing U3 of RNA is kept intact. We were surprised to see Asn become dominant at position 15 after four rounds of selection. It forms a specific base contact with G4 in the wild type U1A-RNA complex. One explanation for this observation is that O3 of U4 was oriented to make a specific contact with the side chain of Asn.

The wild-type U1A binds poorly to RNA sequences containing G4C, G4A, or G4U mutations (*14-16*). The affinity (K_D) of the U1A protein for G4U RNA has been estimated to be approximately 2.1 μM, indicating G4U decreases the affinity of the wild-type U1A by at least 40000-fold. Since Asn15 and Asn16 are selected for binding to G4U RNA, this large decrease in affinity caused by the G4U mutation is of complex origins. A simple perturbation of the binding interface by eliminating hydrogen bonding interactions is not a sufficient explanation for the destabilization observed of the complex formed with G4U RNA. We propose a steric clash may occur in the binding interface of the complex where Glu19 is located upon G4U substitution, possibly between the newly introduced carbonyl group of U4 with the carboxyl group of Glu19. To investigate our hypothesis, we measured the E19A (prepared previously, see Chapter 3) affinity for G4U RNA. The binding energy for the E19A-G4U complex is 8.2 kcal/mol, which is 0.5 kcal/mol higher than the binding energy of the wild type U1A-G4U complex, supporting our hypothesis.

Table 4-2. RNA binding properties of U1A and U1A mutants

Protein	RNA	K_D (M) [a]	ΔG (kcal/mol) [b]	ΔΔG (kcal/mol) [c]
WT	WT	$4.9 \pm 0.3 \times 10^{-11}$	-14.0 ± 0.04	
WT	G4U	$>2.1 \times 10^{-6}$ [d]	<-7.7	6.3
E19Q	WT	$3.1 \pm 0.2 \times 10^{-9}$	-11.6 ± 0.04	
E19Q	G4U	$4.3 \pm 0.1 \times 10^{-8}$	-10.0 ± 0.02	1.6
E19H	WT	$1.2 \pm 0.03 \times 10^{-8}$	-10.8 ± 0.04	
E19H	G4U	$7.5 \pm 0.8 \times 10^{-8}$	-9.7 ± 0.1	1.1
E19A	WT	$2.6 \pm 0.2 \times 10^{-9}$	-11.7 ± 0.04	
E19A	G4U	$8.7 \pm 0.8 \times 10^{-7}$	-8.2 ± 0.1	3.5

[a] K_D is the mean of at least three separate determinations. The standard deviation of these measurements is given.[b] ΔG is the free binding energy of the protein in complex with RNA calculated by the equation $\Delta G = -RT\ln K_D$, where T=298K. [c] ΔΔG is the change in free binding energy upon G4U mutation. [d] We were not able to observe enough bands in the bound form in gels due to the low affinity to evaluate a full binding curve. The affinity of WT-G4U was estimated as the concentration of the protein in the sample containing bound protein. The affinities of the wild type U1A and E19A to the wild type RNA were determined previously (see Chapter 3).

The side chain of Glu19 also makes a specific base contact with U2 in the wild-type U1A-RNA complex. This contact should be maintained and may be responsible for positioning Glu19 close to its original orientation and thus, clashing with the carbonyl group of U4. This spatial clash could be worsened by the limited flexibility of U4 due to its possible contact with Asn15. Compared to the affinity of the E19A U1A mutant for G4U RNA, the substitution of Q or H at position 19 increases the binding energy of the complex by 1.8 or 1.5 kcal/mol. This data is consistent with the formation of two nonlinear hydrogen bonds by Gln or His at position 19 in the U1A mutants. These two residues have a similar arrangement of hydrogen bond donors and acceptors, even though they have different side-chain structures. We propose these two hydrogen bonds in the newly formed binding interface mimic two original ones formed between Glu19 and two bases, U2 and G4, in the wild type complex (Figure 4-1). However, the orientations have

not been optimized, which may result in the weaker strength of these two hydrogen bonds than those (2.3 kcal/mol for E19A mutation) in the wild type complex (see Table 3-1 in Chapter 3). Structural characterization of E19Q or E19H U1A mutants in complex with G4U RNA will be necessary to provide evidence to these hypotheses. We attempted to increase the stability of the E19Q-G4U complex by randomizing the two flanking residues on the each side of Gln19 (positions 17, 18, 20 and 21) in the hope of optimizing the two hydrogen bonding interactions mentioned above. With more stringent selection conditions, no higher affinity binders were isolated after 5 rounds of selection.

We were not able to attain comparable affinities of U1A mutants (E19Q and E19H) for G4U as is observed for the wild type U1A for the cognate RNA. Upon G4U mutation, we lose most of the contacts originally made by G4 due to smaller size and limited number of hydrogen bond donors and acceptors on U4. To determine if these selected U1A mutants can bind to G4U specifically, we measured the affinities of two mutants to the cognate wild-type RNA (Figure 4-5 and Table 4-2). Both the E19Q and E19H U1A mutants show higher affinity for the wild type RNA than for the G4U RNA mutant. E19Q binds to the cognate RNA with a K_D of 3.1 nM, it binds to G4U with a K_D of 43 nM. E19H binds to the cognate RNA with a K_D of 12 nM, lower than E19Q. Its affinity for G4U is shown by a K_D of 75 nM. Although E19Q or E19H are not able to reverse the specificity of the mutant RNA versus the wild-type RNA, interestingly, their presence at position 19 significantly lessens the binding specificity of the wild-type U1A for G4 (see the last column in Table 4-2). For example, the wild-type U1A binds to the cognate RNA with 14 kcal/mol of binding energy. It binds to the G4U RNA mutant with 7.7 kcal/mol of binding energy, suggesting the wild-type U1A strongly prefers to bind to the cognate

RNA. A substitution at position 19 by His shrinks this preference from 6.3 kcal/mol to 1.1 kcal/mol. To reverse the specificity of these mutants, another base will be necessarily mutated, possibly at A1, U3 or U4 positions. Interactions between these bases and the wild-type U1A involve at least one side chain of the protein. Since these bases are separately recognized by discrete segments of the U1A protein, it is possible to isolate binding proteins to designated RNA mutations independently. Those selected residues favorable to mutated RNA binding could be combined to recognize new RNA sequences with higher specificity than the wild type protein.

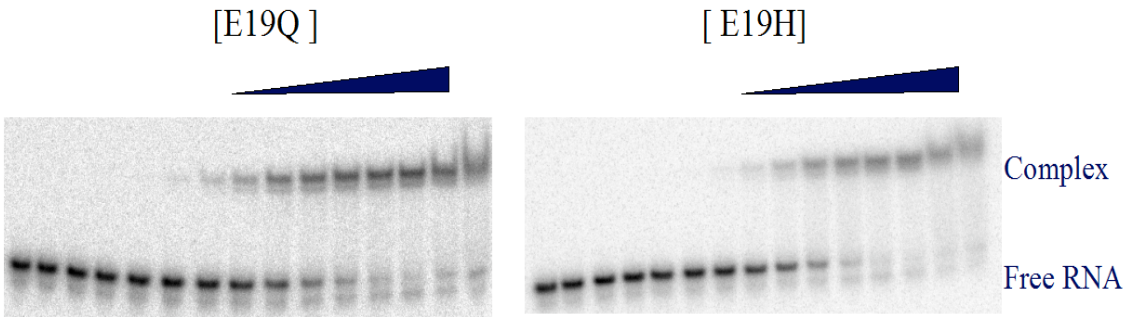


Figure 4-5. Representative electrophoretic mobility shift assays of the E19Q and E19H U1A mutants in complex with the wild type RNA. Samples from a 4-fold serial dilution of the protein were incubated with ^{32}P-labeled RNA at room temperature for 30 min. The shifted bands were resolved by electrophoresis on an 8% native polyacrylamide gel. Control reactions without added protein were loaded in the first lane on the left side of each gel.

4.3 Conclusions

In this study, we demonstrate that Glu at position 19 in the human U1A protein is a key residue for the specific recognition of G4 in the cognate RNA. Selective substitutions at position 19 can alter U1A to bind to a non-cognate G4U RNA mutant with much higher affinity than does the wild-type U1A protein. Asn15 and Asn16, although they are involved in the recognition of G4, are less crucial for specific recognition of G. Whether we can apply the same strategy to select protein binders for other bases in the recognition

element of RNA remains to be tested. With our phage display protocols, we can easily construct phage libraries at approximate 10^7 (the number of individual clones) to achieve complete randomization within 5 target sites. We have also modified the Kunkel method to allow us to randomize sites distant from each other in the primary sequence of the U1A protein. All of these facts provide important tools for designing new U1A-like proteins with more desirable binding properties.

4.4 Materials and methods

4.4.1 Phage display selection *in vitro*

Biopanning experiments were performed as described in Chapter 2. Briefly, ≈ 5.0 X 10^{11} phage were mixed gently with ≈ 2 fmole 5'- biotinylated RNA bearing a G4 mutation (G4A, G4C or G4U) in 30 µl binding buffer (10 mM Tris, 150 mM NaCl, 0.1 % BSA, 0.5% Tween-20, 50 µg of yeast tRNA and 20 units of Rnasin, pH 7.5) for 30 min at room temperature. An 8 µl aliquot of pre-washed streptavidin-coated beads was then added to the binding reaction, and the mixture was incubated for an additional 30 min with shaking. The beads were pelleted by an applied magnetic field, and washed 2-5 times with 1 ml cold wash buffer (10 mM Tris, 150 mM NaCl, 0.1 % BSA, 0.5% Tween-20, pH 7.5). Bound phage were eluted by incubation with 50 µl corresponding unlabeled RNA for 1 hr at room temperature. The eluate was used to infect the log phase XL1-Blue cells as described in Chapter 2. For the fourth round of selection, only 0.2 fmole 5'-biotinylated G4U RNA was applied to select RNA-binding phage with higher affinity.

4.4.2 Phage-ELISA

To validate the binding properties of the selected clones for the G4U RNA mutant from the U1A phage library, we performed phage ELISA. All purified phage solutions

were prepared from single well-isolated clones of *E. coli* XL1-B to avoid cross-contamination. Wells on a streptavidin-coated plate (Pierce) were coated with 50 µl 5'-biotinylated G4U RNA at 5 µg/ml in wash buffer (10 mM Tris, 150 mM NaCl, 0.1 % BSA, 0.5 % Tween-20, pH 7.5) for 1 hr at room temperature with gentle shaking. After washing three times with 200 µl wash buffer, 100 µl of phage solutions (1.0 X 10^{12} phage/ml) were added to the designated wells. We applied 100 µl wash buffer to one well and the same concentration of helper phage to another as controls. After incubation for an additional 30 min, unbound phage were washed off. Anti-M13/horseradish peroxidase conjugate (100 µl, 1:5000 in wash buffer) was then added to wells and incubated for 1 hr. After three washes, assays were developed with 1-Step Ultra TMB-ELISA (Pierce) for 15 min at room temperature. Reactions were stopped with 100 µl 2 M sulfuric acid and read spectrophotometrically at 450 nm.

4.4.3 Protein mutagenesis and purification

The expression vectors for U1A mutants E19Q, E19H were constructed by the Kunkel method (Figure 4-6). Proteins were expressed as His_6-tagged fusions in *E. coli* strain BL21(DE3)pLysS and purified (Figures 4-7, 4-8 and 4-9). These protocols are described in detail in Chapter 3.

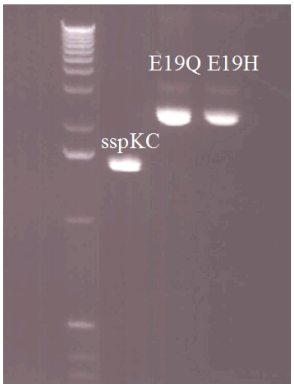


Figure 4-6. Enzymatic synthesis of covalently closed circular, double-stranded DNA containing E19Q and E19H mutations. The eluted heteroduplex DNA was electrophoresed on a 1.0 % TBE/agarose gel containing ethidium bromide for DNA visualization.

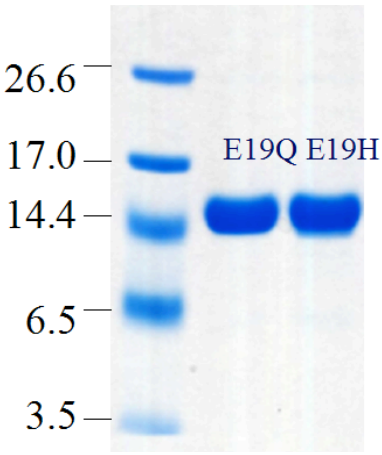


Figure 4-7. Analysis of the purity of E19Q and E19H U1A mutants by 12 % SDS-PAGE. Molecular weight markers are in kDa

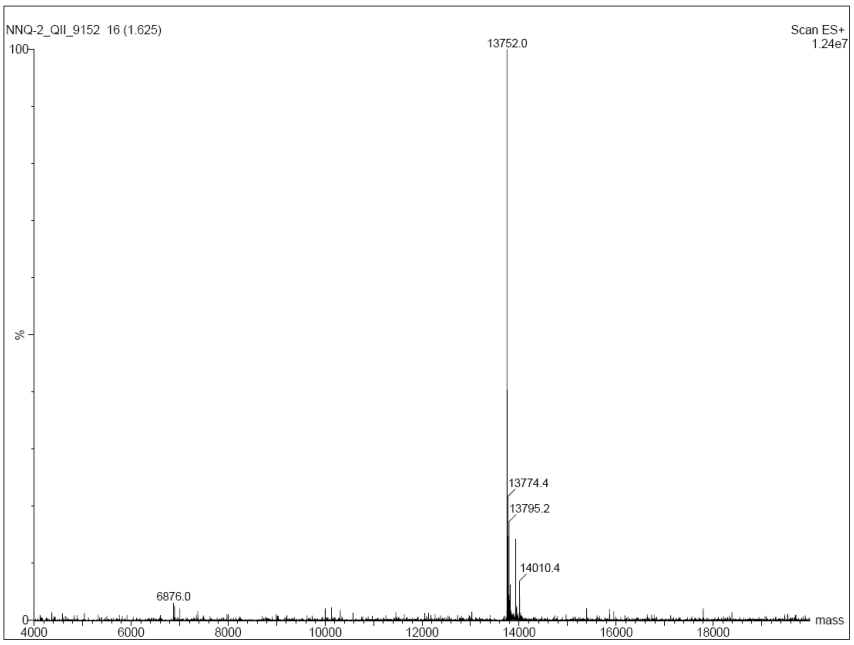


Figure 4-8. Electrospray ionization mass spectrum of E19Q U1A mutant. Calculated MW: 13754 Daltons. Observed MW: 13752 Daltons.

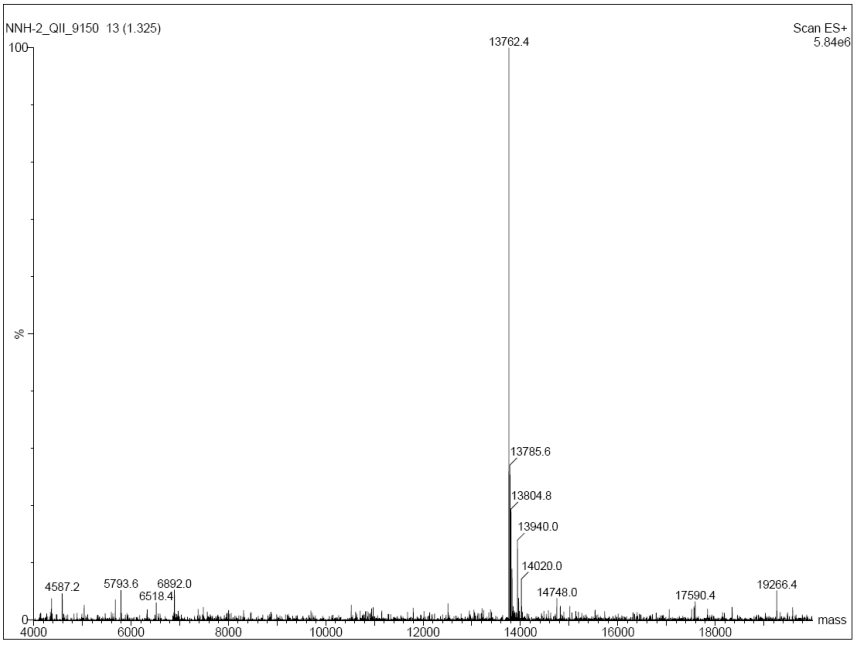


Figure 4-9. Electrospray ionization mass spectrum of E19Q U1A mutant. Calculated MW: 13764 Daltons. Observed MW: 13762 Daltons.

4.4.4 Binding affinity assays

Electrophoretic mobility shift assays (EMSA) were used to determine the binding constants of proteins to RNA as described in Chapter 3. The affinities of the wild type U1A and E19A to the wild type RNA were determined previously (see Chapter 3).

4.5 References

1. Caban, K., Kinzy, S. A., and Copeland, P. R. (2007) The L7Ae RNA binding motif is a multifunctional domain required for the ribosome-dependent Sec incorporation activity of Sec insertion sequence binding protein 2, *Mol Cell Biol 27*, 6350-6360.

2. Cheng, S., and Gallie, D. R. (2007) eIF4G, eIFiso4G, and eIF4B bind the poly(A)-binding protein through overlapping sites within the RNA recognition motif domains, *J Biol Chem 282*, 25247-25258.

3. Lin, J. C., and Tarn, W. Y. (2009) RNA binding motif protein 4 translocates to cytoplasmic granules and suppresses translation via Argonaute2 during muscle cell differentiation, *J Biol Chem*.

4. Maris, C., Dominguez, C., and Allain, F. H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression, *FEBS J 272*, 2118-2131.

5. Uranishi, H., Zolotukhin, A. S., Lindtner, S., Warming, S., Zhang, G. M., Bear, J., Copeland, N. G., Jenkins, N. A., Pavlakis, G. N., and Felber, B. K. (2009) The RNA-binding motif protein 15B (RBM15B/OTT3) acts as cofactor of the nuclear export receptor NXF1, *J Biol Chem 284*, 26106-26116.

6. Wang, S., Hu, Y., Overgaard, M. T., Karginov, F. V., Uhlenbeck, O. C., and McKay, D. B. (2006) The domain of the Bacillus subtilis DEAD-box helicase YxiN that is responsible for specific binding of 23S rRNA has an RNA recognition motif fold, *RNA 12*, 959-967.

7. Lunde, B. M., Moore, C., and Varani, G. (2007) RNA-binding proteins: modular design for efficient function, *Nat Rev Mol Cell Biol 8*, 479-490.

8. Finn, R. D., Tate, J., Mistry, J., Coggill, P. C., Sammut, S. J., Hotz, H. R., Ceric, G., Forslund, K., Eddy, S. R., Sonnhammer, E. L., and Bateman, A. (2008) The Pfam protein families database, *Nucleic Acids Res 36*, D281-288.

9. Dreyfuss, G., Swanson, M. S., and Pinol-Roma, S. (1988) Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation, *Trends Biochem Sci 13*, 86-91.

10. Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K., and Dreyfuss, G. (1986) mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence, *Mol Cell Biol 6*, 2932-2943.

11. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994) Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin, *Nature 372*, 432-438.

12. Boelens, W. C., Jansen, E. J., van Venrooij, W. J., Stripecke, R., Mattaj, I. W., and Gunderson, S. I. (1993) The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA, *Cell 72*, 881-892.

13. van Gelder, C. W., Gunderson, S. I., Jansen, E. J., Boelens, W. C., Polycarpou-Schwarz, M., Mattaj, I. W., and van Venrooij, W. J. (1993) A complex secondary structure in U1A pre-mRNA that binds two molecules of U1A protein is required for regulation of polyadenylation, *EMBO J 12*, 5191-5200.

14. Zeng, Q., and Hall, K. B. (1997) Contribution of the C-terminal tail of U1A RBD1 to RNA recognition and protein stability, *RNA 3*, 303-314.

15. Showalter, S. A., and Hall, K. B. (2004) Altering the RNA-binding mode of the U1A RBD1 protein, *J Mol Biol 335*, 465-480.

16. Benitex, Y., and Baranger, A. M. (2007) Recognition of essential purines by the U1A protein, *BMC Biochem 8*, 22.

17. Greisman, H. A., and Pabo, C. O. (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites, *Science 275*, 657-661.

18. Belien, T., Van Campenhout, S., Vanden Bosch, A., Bourgois, T. M., Rombouts, S., Robben, J., Courtin, C. M., Delcour, J. A., and Volckaert, G. (2007) Engineering molecular recognition of endoxylanase enzymes and their inhibitors through phage display, *J Mol Recognit 20*, 103-112.

19. Sidhu, S. S., and Koide, S. (2007) Phage display for engineering and analyzing protein interaction interfaces, *Curr Opin Struct Biol 17*, 481-487.

20. Simon, M. D., Feldman, M. E., Rauh, D., Maris, A. E., Wemmer, D. E., and Shokat, K. M. (2006) Structure and properties of a re-engineered homeodomain protein-DNA interface, *ACS Chem Biol 1*, 755-760.

21. Cheng, A. C., Calabro, V., and Frankel, A. D. (2001) Design of RNA-binding proteins and ligands, *Curr Opin Struct Biol 11*, 478-484.

22. Cheong, C. G., and Hall, T. M. (2006) Engineering RNA sequence specificity of Pumilio repeats, *Proc Natl Acad Sci U S A 103*, 13635-13639.

23. Laird-Offringa, I. A., and Belasco, J. G. (1996) In vitro genetic analysis of RNA-binding proteins using phage display libraries, *Methods Enzymol 267*, 149-168.

CHAPTER 5

IDENTIFICATION OF DOMAINS OF CUG-BP1 REQUIRED FOR GRE BINDING

5.1 Introduction

CUG-binding protein 1 (CUG-BP1) was first identified as a protein that binds to expanded CUG repeats in the 3'- untranslated region (3'-UTR) of dystrophia myotonia-protein kinase (DMPK) gene (*1-2*). The elevated expression level of this protein is thought to contribute to muscle weakness, myotonia and insulin resistance of myotonic dystrophy type 1 (DM1) (*3-5*). CUG-BP1 participates in the regulation of post-transcriptional processing of mRNA. For example, in chickens and humans, CUG-BP1 has been shown to bind to GU rich repeats in the cardiac troponin T (cTNT) gene to regulate MSE (muscle-specific splicing enhancer)-dependent exon inclusion during pre-mRNA splicing (*6-8*). In the *Xenopus* oocyte, the EDEN-binding protein (EDEN-BP, the *X. laevis* ortholog of CUP-BP1) acts as a *trans*-acting factor to accelerate the deadenylation of the Eg5 mRNA by binding to a UR rich repeat in the embryo deadenylation element (EDEN) (*9*). CUG-BP1 has also been reported to bind to a GU rich region in both C/EBPβ and p21 mRNAs to enhance mRNA translation and stability (*10*).

Recently, a new function of CUG-BP1 has been revealed in which CUG-BP1 acts as a functional mRNA decay-promoting RNA binding protein by recognizing a GU rich element (GRE) in the 3'-UTR of transcripts in primary human T cells (*11*). As the best characterized example of selective mRNA decay mechanisms, AU rich elements (AREs) were found to be conserved in a variety of short-lived transcripts for encoding transient

proteins required for inflammatory responses, stress response, cell division and proliferation (*12-13*). Many ARE-binding proteins (ARE-BPs) have been identified to participate in the ARE-directed rapid mRNA decay and have provided insight into the functional role of these RNA-binding proteins in RNA metabolism and gene expression. In their recent study, Bohjanen and coworkers found many transcripts that showed rapid decay in activated T cells, but did not contain AREs or other known RNA regulatory elements. They identified a new consensus decay-signaling sequence UGUUUGUUUGU in a 3'-untranslated region of these transcripts. Through supershift and immunoprecipitation assays, CUG-BP1 was shown to be the major binding protein interacting with the GRE in cytoplasmic extracts from primary human T cells. Knockdown of CUB-BP1 by RNA interference stabilized the mRNA transcripts containing GREs, suggesting GRE-dependent mRNA decay is mediated by the association with CUG-BP1 (*11*). These findings reveal a novel mRNA decay mechanism and provide more insight into understanding of regulated gene expression by a conserved pathway that selectively mediates mRNA decay.

CUG-BP1 is a member of Bruno-like protein family (*8, 14-15*). It has a unique domain structure comprised of three RNA recognition motifs (RRMs): two closely spaced RRMs (RRM1 and RRM2) near the N-terminus and a single RRM (RRM3) at the C-terminus, linked by a 208-residue divergent domain (Figure 5-1) (*8*). Besides the GRE element described above, at least three other target RNA sequences, namely CUG, UG and UA repeats, have been identified as RNA targets of CUG-BP1 (*16-17*). In different biological contexts, the three RRMs of CUG-BP1 exhibit distinct binding preferences for RNA targets by using various combinations of these RRM modules. For example, the

two consecutive RRMs (RRM1 and RRM2) at the N-terminus of the protein cooperatively recognize CUG repeats, but RRM1 and RRM2 individually have no affinity for CUG repeats (*16-17*). RRM3 itself also does not bind to CUG repeats, but it does bind to UG repeats (*17*). Either RRM1 or RRM2 of CELF4, another Bruno-like protein, is able to bind to UG rich element within MSEs. A peptide comprised of RRM2 and 66 residues of the divergent linker shows efficient MSE-dependent splicing *in vivo* that is comparable to the full-length protein (*18*). In this study, we have performed a deletion analysis of CUGBP1 to define the contributions of individual RRMs and RRM combinations to the high stability of the complex formed between CUG-BP1 and the GRE sequence.

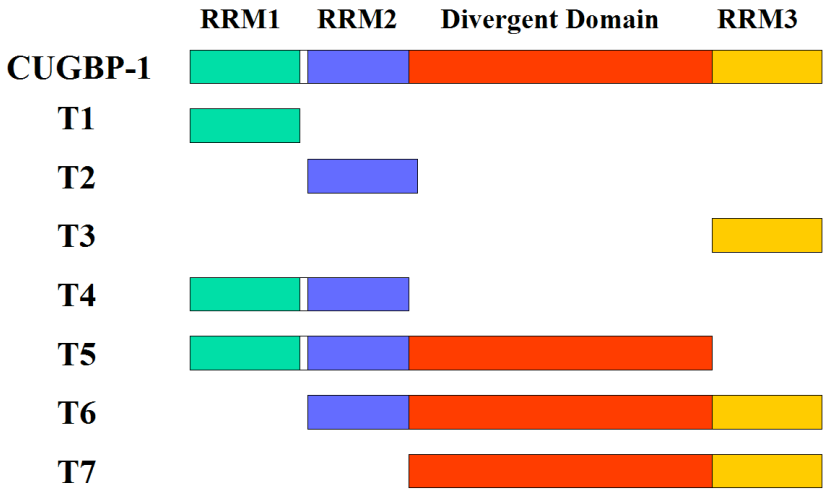


Figure 5-1. Schematic diagram of deletion mutants of CUG-BP1.

5.2 Preliminary results

CUG-BP1 has been shown to act as a *trans*-acting factor that mediates GRE-dependent mRNA decay in primary human T cells. It binds specifically to GRE sequences with high affinity (*11*). Although the three RRMs of CUG-BP1 maintain the same structural topology as other RRMs, they have little sequence similarity (*8*), suggesting they may contribute differently to the affinity of CUG-BP1 for GRE sequences. RRM1 and RRM2 have been shown to efficiently bind to a UG repeat in CELF4 through a yeast three-hybrid assay (*18*).

Several deletion mutants of CUG-BP1 have been constructed (Figure 5-1) and their individual contributions to the affinity for the GRE sequence were analyzed. These proteins were overexpressed in *E. coli* cells with a histidine tag fused at the N-terminus. We were not able to obtain pure full-length CUG-BP1 (Figure 5-2). CUG-BP1 is 480 residues long, and the impurities below the major band could be due to the premature termination of transcription or translation. This phenomenon has been observed previously with the purification of the N-terminal His-tagged CUG-BP1 in another expression vector (pTRC) (*19*). All of the deletion mutants of CUG-BP1 that contain only the RRM sequences (T1-T4) were expressed well and purified to > 95% (Figure 5-3, lane 2-5). T6 and T7, which contain the divergent domain and RRM3, failed to express (Figure 5-3, lane 6 and 8). T5, which is comprised of RRM1, RRM2, and the divergent domain, showed multiple bands on SDS-PAGE after freshly eluted from the Ni^{2+} column, it is subjected to proteolytic cleavage (Figure 5-3, lane 7). Interestingly, the size of the cleaved form is comparable to that of T4. All these observations suggest RRM1 and RRM2 may be important for maintaining the stability of the full-length protein.

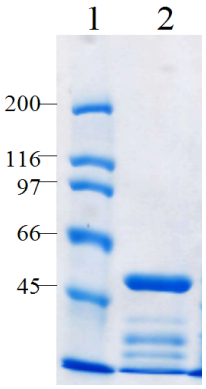


Figure 5-2. Coomassie blue staining of the purified recombinant CUG-BP1.
Lane 1: molecular weight markers in kDa.
Lane 2: the recombinant CUG-BP1.

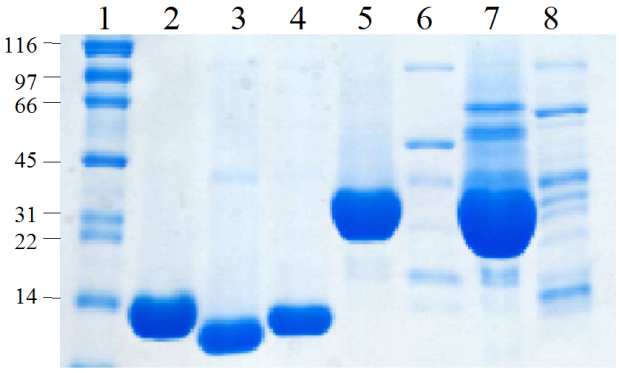


Figure 5-3. Coomassie blue staining of the purified deletion mutants of CUG-BP1.

Lane 1: molecular weight markers in kDa. Lane 2: T1. Lane 3: T2. Lane 4: T3. Lane 5: T4. Lane 6: T6. Lane 7: T5. Lane 8: T7.

To determine whether these purified fragments of CUG-BP1 (T1, T2, T3 and T4) could bind to the GRE sequence, we analyzed their affinities to the RNA target by EMSA (Figure 5-4, Table 5-1). Two complexes were formed with increasing concentrations of deletion mutants, consistent with previously reported observations (*19-21*). The lanes for free RNA showed two distinct bands, possibly due to heterogeneity in the RNA population. The full-length CUG-BP1 showed high affinity for GRE as shown by multiple bands of the bound complexes; the K_D was estimated to approximately 1.6 nM. RRM1 (T1) itself only forms the complex with GRE at the highest concentration of recombinant protein applied. The K_D was estimated to be approximately 150 μM. RRM 2

(T2) showed much higher affinity for GRE than RRM1, with a K_D of 8.6 μM. The peptide comprised of RRM1 and RRM2 (T4) binds with similar affinity for GRE as RRM2, suggesting RRM1 is not require for GRE recognition. RRM1 of CUG-BP1 has been shown to be important for the recognition of UG repeats and CUG repeats (*16*-*17*). The K_D of RRM3 for GRE is 0.87 μM, the smallest among all three RRMs evaluated. This value is similar to the recently reported K_D of RRM3 for a short UG repeat, suggesting the binding mechanism of recognition of UG repeats and GRE may be similar. Among the three RRMs of CUG-BP1, the sequence of RRM3 is highly conserved within all Bruno-like proteins (*15*). It is likely that RRM3 and RRM2 cooperatively bind to the GRE sequence, while RRM3 contributes the majority of binding energy to the complex between CUG-BP1 and GRE. Further investigation of RNA binding by RRM3 itself will be important for elucidating the functions of CUG-BP1 and other Bruno-like proteins. The crystal structure of human CUG-BP1 RRM3 bound to a UG repeat (UGUGUG) has appeared recently (*22*).

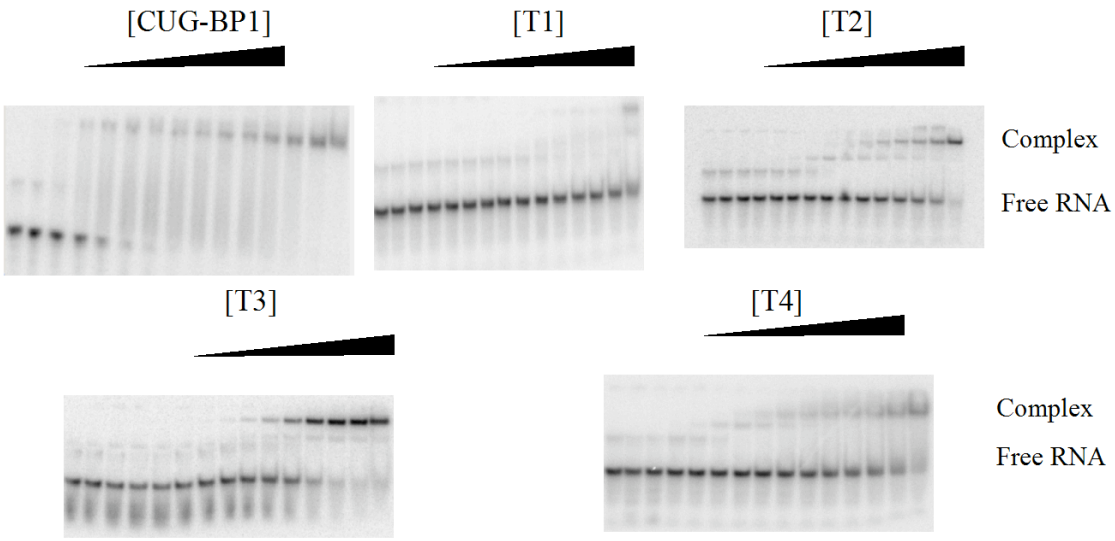


Figure 5-4. Representative electrophoretic mobility shift assays of CUG-BP1 and its deletion mutants in complex with the GRE sequence. Samples from a 3-fold serial dilution of the designated protein were incubated with ^{32}P-labeled GRE at room temperature for 30 min. The shifted bands were resolved by electrophoresis on an 8% native polyacrylamide gel (a 6 % native polyacrylamide gel was used for the full-length CUG-BP1). Control reactions without added protein were loaded in the first lane on the left side of each gel.

Table 5-1. GRE binding properties of CUG-BP1 and its deletion mutants

Protein	K_D (M) [a]	ΔG (kcal/mol) [b]
WT	$<1.6 \times 10^{-9}$ [c]	>-12.0
T1	$>1.5 \times 10^{-4}$ [d]	<-5.2
T2	$8.6 \pm 0.8 \times 10^{-6}$	-6.9 ± 0.05
T3	$8.7 \pm 0.7 \times 10^{-7}$	-8.2 ± 0.04
T4	$6.7 \pm 0.8 \times 10^{-6}$	-7.0 ± 0.07

[a] K_D is the mean of at least three separate determinations. The standard deviation of these measurements is given.[b] ΔG is the free binding energy of the protein in complex with RNA calculated by the equation $\Delta G = -RT\ln K_D$, where T=298K. [c] We cannot determine the concentration of CUG-BP1 accurately due to the impurities. [d] We were not able to observe enough bands in the bound form to evaluate a full binding curve due to the low affinity. The affinity was estimated as the concentration of the protein in the sample containing bound protein.

5.3 Materials and methods

5.3.1 Expression vectors

Individual deletion mutants of CUG-BP1 were defined as follows: T1 (Met1-Ser99), T2 (Asp107-Thr189), T3 (Ala398-Asp480), T4 (Met1-Thr189), T5 (Met1-Gly397), T6 (Asp107-Asp480), T7 (Gln190-Asp480). The DNA fragments encoding these deletion mutants were amplified by PCR (Figure 5-5) from the full-length cDNA clone in pET-15b (obtained from Professor Swanson at University of Florida) and cloned back into the expression vector pET-15b between the *BamHI* and *NdeI* sites. pET-15b contains a His affinity tag upstream of the *BamHI* site. All cloned DNA sequences were confirmed by sequencing.

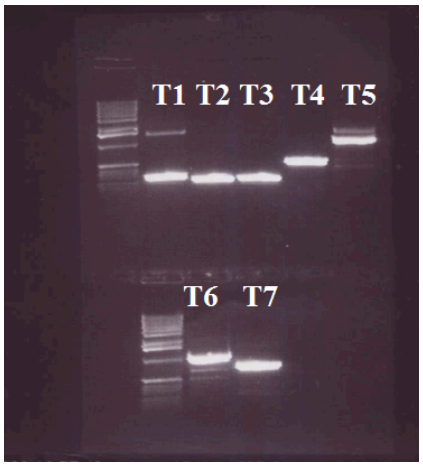


Figure 5-5. PCR of DNA fragments encoding deletion mutants of CUG-BP1. The amplified DNA was electrophoresed on a 1.0% TBE/agarose gel containing ethidium bromide for DNA visualization.

The forward and reverse primers for PCR were purchased from IDT and purified by 12 % denaturing PAGE (see the protocol in Chapter 3). The sequences are listed as follows:

T1: 5'-GGAATTCCATATGATGAACGGCACCCTG-3' (forward)

5'-CGGGATCCTCAACTGTCAGCAGG-3' (reverse)

T2: 5'-GGAATTCCATATGATGGACAGGAAGCTGTTTATTG-3'(forward)

5'-CGGGATCCTCATGTATCAGCAAATTTTACCAC-3'(reverse)

T3: 5’-GGAATTCCATATGATGGCCAACCTGTTCATC-3’ (forward)

5’-CGGGATCCTCAGTAGGGCTTGCTG-3’(reverse)

T4: 5’-GGAATTCCATATGATGAACGGCACCCTG-3’ (forward)

5’-CGGGATCCTCATGTATCAGCAAATTTTACCAC-3’(reverse)

T5: 5’-GGAATTCCATATGATGAACGGCACCCTG-3’ (forward)

5’-CGGGATCCTCATCCCTCTGGACC-3’ (reverse)

T6: 5’-GGAATTCCATATGATGGACAGGAAGCTGTTTATTG-3’(forward)

5’-CGGGATCCTCAGTAGGGCTTGCTG-3’(reverse)

T7: 5’-GGAATTCCATATGATGCAGAAGGACAAAGAACAG-3’ (forward)

5’-CGGGATCCTCAGTAGGGCTTGCTG-3’(reverse)

5.3.2 Protein expression and purification

Small proteins (T1, T2, T3 and T4) were overexpressed as His_6-tagged fusions in *E. coli* strain *Rosetta* 2 and purified as described in Chapter 3. For T5, T6, T7 and the full-length CUG-BP1, a 800 ml culture was used and cells were induced at early log phase ($OD_{600} \approx 0.4$) with 0.5 mM isopropyl 1-thio-P-D-glactopyranoside. The culture was grown for an additional 1.5 hr at 37 ^{0}C and followed the purification protocol in Chapter 3. The concentrations of the proteins were estimated by the Bradford assay.

5.3.3 Binding affinity assays

Electrophoretic mobility shift assays (EMSA) were used to determine the affinities of various deletion mutants of CUG-BP1 for the GRE sequence. The general procedure was described in Chapter 3. The 22-mer synthetic GRE (5’-UUUCUUGUUUGUUUGUUUGGGU-3’) was purchased from IDT and purified on a 10 % denaturing polyacrylamide gel and 5’ end-labeled with [γ-^{32}P] ATP as described. A 6

% native polyacrylamide gel was used for binding assays with CUG-BP1. For all other proteins, 8 % native polyacrylamide gels were used.

5.4 References

1. Timchenko, L. T., Timchenko, N. A., Caskey, C. T., and Roberts, R. (1996) Novel proteins with binding specificity for DNA CTG repeats and RNA CUG repeats: implications for myotonic dystrophy, *Hum Mol Genet 5*, 115-121.

2. Timchenko, L. T., Miller, J. W., Timchenko, N. A., DeVore, D. R., Datar, K. V., Lin, L., Roberts, R., Caskey, C. T., and Swanson, M. S. (1996) Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy, *Nucleic Acids Res 24*, 4407-4414.

3. Philips, A. V., Timchenko, L. T., and Cooper, T. A. (1998) Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy, *Science 280*, 737-741.

4. Savkur, R. S., Philips, A. V., and Cooper, T. A. (2001) Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy, *Nat Genet 29*, 40-47.

5. Kuyumcu-Martinez, N. M., Wang, G. S., and Cooper, T. A. (2007) Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation, *Mol Cell 28*, 68-78.

6. Ryan, K. J., and Cooper, T. A. (1996) Muscle-specific splicing enhancers regulate inclusion of the cardiac troponin T alternative exon in embryonic skeletal muscle, *Mol Cell Biol 16*, 4014-4023.

7. Cooper, T. A. (1998) Muscle-specific splicing of a heterologous exon mediated by a single muscle-specific splicing enhancer from the cardiac troponin T gene, *Mol Cell Biol 18*, 4519-4525.

8. Ladd, A. N., Charlet, N., and Cooper, T. A. (2001) The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing, *Mol Cell Biol 21*, 1285-1296.

9. Paillard, L., Omilli, F., Legagneux, V., Bassez, T., Maniey, D., and Osborne, H. B. (1998) EDEN and EDEN-BP, a cis element and an associated factor that mediate sequence-specific mRNA deadenylation in Xenopus embryos, *EMBO J 17*, 278-287.

10. Salisbury, E., Sakai, K., Schoser, B., Huichalaf, C., Schneider-Gold, C., Nguyen, H., Wang, G. L., Albrecht, J. H., and Timchenko, L. T. (2008) Ectopic expression of cyclin D3 corrects differentiation of DM1 myoblasts through activation of RNA CUG-binding protein, CUGBP1, *Exp Cell Res 314*, 2266-2278.

11. Vlasova, I. A., Tahoe, N. M., Fan, D., Larsson, O., Rattenbacher, B., Sternjohn, J. R., Vasdewani, J., Karypis, G., Reilly, C. S., Bitterman, P. B., and Bohjanen, P. R. (2008) Conserved GU-rich elements mediate mRNA decay by binding to CUG-binding protein 1, *Mol Cell 29*, 263-270.

12. Chen, C. Y., and Shyu, A. B. (1995) AU-rich elements: characterization and importance in mRNA degradation, *Trends Biochem Sci 20*, 465-470.

13. Barreau, C., Paillard, L., and Osborne, H. B. (2005) AU-rich elements and associated factors: are there unifying principles?, *Nucleic Acids Res 33*, 7138-7150.

14. Good, P. J., Chen, Q., Warner, S. J., and Herring, D. C. (2000) A family of human RNA-binding proteins related to the Drosophila Bruno translational regulator, *J Biol Chem 275*, 28583-28592.

15. Barreau, C., Paillard, L., Mereau, A., and Osborne, H. B. (2006) Mammalian CELF/Bruno-like RNA-binding proteins: molecular characteristics and biological functions, *Biochimie 88*, 515-525.

16. Timchenko, L. T. (1999) Myotonic dystrophy: the role of RNA CUG triplet repeats, *Am J Hum Genet 64*, 360-364.

17. Mori, D., Sasagawa, N., Kino, Y., and Ishiura, S. (2008) Quantitative analysis of CUG-BP1 binding to RNA repeats, *J Biochem 143*, 377-383.

18. Singh, G., Charlet, B. N., Han, J., and Cooper, T. A. (2004) ETR-3 and CELF4 protein domains required for RNA binding and splicing activity in vivo, *Nucleic Acids Res 32*, 1232-1241.

19. Marquis, J., Paillard, L., Audic, Y., Cosson, B., Danos, O., Le Bec, C., and Osborne, H. B. (2006) CUG-BP1/CELF1 requires UGU-rich sequences for high-affinity binding, *Biochem J 400*, 291-301.

20. Cosson, B., Gautier-Courteille, C., Maniey, D., Ait-Ahmed, O., Lesimple, M., Osborne, H. B., and Paillard, L. (2006) Oligomerization of EDEN-BP is required for specific mRNA deadenylation and binding, *Biol Cell 98*, 653-665.

21. Graindorge, A., Le Tonqueze, O., Thuret, R., Pollet, N., Osborne, H. B., and Audic, Y. (2008) Identification of CUG-BP1/EDEN-BP target mRNAs in Xenopus tropicalis, *Nucleic Acids Res 36*, 1861-1870.

22. Tsuda, K., Kuwasako, K., Takahashi, M., Someya, T., Inoue, M., Terada, T., Kobayashi, N., Shirouzu, M., Kigawa, T., Tanaka, A., Sugano, S., Guntert, P., Muto, Y., and Yokoyama, S. (2009) Structural basis for the sequence-specific RNA-recognition mechanism of human CUG-BP1 RRM3, *Nucleic Acids Res 37*, 5151-5166.